

Vlad BOCĂNEȚ

Statistică și probabilități



Îndrumător de laborator

**Editura UTPRESS
Cluj Napoca, 2023
ISBN 978-606-737-672-2**

Vlad BOCĂNEȚ

Statistică și probabilități

Îndrumător de laborator



UTPRESS
Cluj-Napoca, 2023
ISBN 978-606-737-672-2



Editura UTPRESS
Str. Observatorului nr. 34
400775 Cluj-Napoca
Tel.: 0264-401.999
e-mail: utpress@biblio.utcluj.ro
www.utcluj.ro/editura

Director: ing. Dan COLȚEA

Recenzia: Prof.dr. Ioan Cristian Chifu
Prof.dr.ing. Marius Bulgaru

Pregătire format electronic on-line: Gabriela Groza

Copyright © 2023 Editura UTPRESS
Reproducerea integrală sau parțială a textului sau ilustrațiilor
din această carte este posibilă numai cu acordul prealabil
scris al editurii UTPRESS.

ISBN 978-606-737-672-2

Cuprins

Introducere.....	6
1. Colectarea și procesarea primară a datelor.....	8
1.1. Obiective de învățare.....	8
1.2. Concepte teoretice.....	8
1.2.1. Generarea unui șir de numere aleatorii.....	10
1.2.2. Noțiuni generale privind folosirea funcțiilor în Excel.....	12
1.2.3. Sortarea valorilor.....	13
1.2.4. Importarea datelor din surse externe.....	15
1.2.5. Curățarea datelor.....	17
1.2.6. Filtrarea datelor.....	18
1.2.7. Trasarea unei diagrame.....	20
1.3. Descrierea temei.....	22
1.4. Resurse.....	28
2. Determinarea indicatorilor statistici de localizare și împrăștiere și reprezentarea grafică a datelor.....	32
2.1. Obiective de învățare.....	32
2.2. Concepte teoretice.....	32
2.2.1. Indicatori de localizare.....	33
2.2.2. Indicatori de împrăștiere.....	36
2.2.3. Funcții Excel pentru calculul indicatorilor statistici.....	37
2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvențe....	40
2.2.5. Trasarea histogramei.....	43
2.3. Descrierea temei.....	46
2.4. Resurse.....	53

3.	Eliminarea valorilor aberante și determinarea coeficientului de rebut	56
3.1.	Obiective de învățare.....	56
3.2.	Concepte teoretice	56
3.2.1.	Eșantionarea.....	56
3.2.2.	Valorile aberante.....	58
3.2.3.	Determinarea coeficientului de rebut	60
3.3.	Descrierea temei.....	63
3.4.	Resurse	68
4.	Estimarea parametrilor populației.....	70
4.1.	Obiective de învățare.....	70
4.2.	Concepte teoretice	70
4.2.1.	Estimarea mediei.....	73
4.2.2.	Estimarea dispersiei	87
4.3.	Descrierea temei.....	94
4.4.	Resurse	98
5.	Determinarea capabilității și crearea unei cartele de control	99
5.1.	Obiective de învățare	99
5.2.	Concepte teoretice.....	99
5.2.1.	Capabilitate	105
5.2.2.	Structura generală a cartelelor de control a calității	106
5.2.3.	Corelația	109
5.2.4.	Regresia liniară	111
5.3.	Descrierea temei	113
5.4.	Resurse	115
	Anexa 1 – Exemple de rapoarte	116
	Anexa 2 - Tabel pentru distribuția normal (valori z)	123

Anexa 3 - Tabelul distribuției Student (valori t)	124
Anexa 4 - Tabelul distribuției Chi-pătrat (valori χ^2)	125
Tabel de figuri.....	126
Tabel de tabele.....	127

Introducere

Acest îndrumător de lucrări de laborator se adresează în special studenților de anul II de la specializările Inginerie Industrială și Inginerie Economică Industrială. Lucrările prezentate își propun o abordare practică a unor probleme întâlnite tot mai des de ingineri în munca lor curentă. Se face o introducere graduală în problemele de culegere, curățare, procesare, vizualizare și interpretare a datelor generate de procese tehnologice și din domeniul industriei.

Deoarece lucrările din acest îndrumător au rolul de punere în practică a cunoștințelor acumulate, în secțiunea de *Concepte teoretice*, prezentă în fiecare lucrare, nu se vor aprofunda conceptele teoretice prezentate la curs ci se va face o reamintire a acestora și se va explica modul în care acestea se aplică la problema tratată.

Lucrările de laborator folosesc ca unealtă principală MS Excel (versiunea 365). Informațiile sunt astfel structurate încât să permită rezolvarea problemelor propuse cu ajutorul Excel-ului, dar bineînțeles că acestea pot fi rezolvate și cu alte unelte de calcul tabelar sau de analiză statistică. Varianta de Excel folosită este cea în limba engleză, aceasta fiind versiunea cea mai comună. Pentru versiunile în alte limbi va trebui să identificați butoanele/funțiile echivalente celor din limba engleză.

În cadrul acestui îndrumător sunt introduse link-uri la diferite resurse externe menite a vă ajuta să consolidați informația învățată sau pentru aprofundarea subiectului tratat. La momentul scrierii acestei lucrări toate link-urile sunt accesibile și relevante dar autorii nu pot garanta că acest lucru va fi valabil și la momentul citirii. Drept urmare, dacă din motive independente de autori, aceste resurse devin inaccesibile, vă rugăm să efectuați căutări pe internet folosind orice motor de căutare pentru găsirea resurselor relevante pentru subiectul pe care doriți să-l aprofundați.

Prima lucrare se concentrează pe laborioasa componentă a achiziției și pre-procesare a datelor primare. Studentul va învăța diversele tipuri de date, surse de date, cum se importă sau se generează astfel de date și o primă procesare a acestora prin sortarea și filtrarea lor. De asemenea va învăța cum se curăță datele, eliminând valorile duplicate și sau cele invalide și va face validarea datelor în funcție de formatul necesar.

A doua lucrare introduce elemente de bază de analiză statistică precum indicatorii cei mai folosiți de localizare și împrăștiere. Studentul va învăța ce este și cum se calculează media, mediana și modala, va face cunoștință cu noțiunile de frecvență, dispersie și abatere standard și va învăța diferența dintre diferiții indicatori.

În cea de-a treia lucrare studentul va face cunoștință cu valorile aberante și va învăța metoda cuartilelor pentru eliminarea acestora. După aplicarea acestei metode pe un set de date, va determina grafic coeficientul de rebut pentru un set de măsurători folosind o histogramă.

A patra lucrare de laborator face o trecere prin diferite scenarii de estimare a parametrilor unei populații statistice (media și dispersia) cu toate tipurile de risc. Studentul învață să folosească tabelele de distribuție pentru distribuția Normală, Student și Chi-pătrat.

Ultima lucrare abordează instrumentele SPC: Aici studentul va învăța să creeze o cartelă de control, va exersa crearea histogramei și va determina capabilitatea unui proces. Va exersa de asemenea calculul corelației și coeficientului de determinare precum și reprezentarea grafică a unei drepte de regresie.

1. Colectarea și procesarea primară a datelor

1.1. [Obiective de învățare](#)

1.2. [Concepte teoretice](#)

1.3. [Descrierea temei](#)

1.4. [Resurse](#)

1.1. Obiective de învățare

În această lucrare veți învăța diferite moduri de generare sau colectare de date. Datele sunt esențiale pentru realizarea de analize statistice. Un factor foarte important este structura și corectitudinea datelor colectate. Pentru a obține rezultate corecte, datele nu trebuie să conțină erori și drept urmare trebuie „curățate” înainte de a fi analizate. Metode precum filtrarea și sortarea datelor ne pot da informații despre corectitudinea acestora și ne pot ajuta în curățarea datelor.

După terminarea acestei lucrări veți putea să:

- Faceți distincție între diferite tipuri de date;
- Generați un șir de date aleatorii după o anumită distribuție;
- Importați date într-un registru Excel din surse externe;
- Sortați și filtrați valorile dintr-un registru;
- Identificați și să eliminați valorile duplicate sau eronate și a intrărilor cu valori lipsă;
- Sintetizați informațiile obținute într-un Raport.

1.2. Concepte teoretice

Tehnologiile actuale ne permit să culegem din ce în ce mai multe date din mediul nostru înconjurător. Aceste date provin dintr-o multitudine de surse și pot fi interpretate în diverse moduri. Identificarea corectă a tipului de date și

folosirea corespunzătoare este esențială în interpretarea semnificației acestora și a înțelegerii fenomenului pe care îl studiem.

Din punct de vedere statistic datele pot fi *cantitative* sau *calitative*. Datele cantitative sunt asociate mărimilor care pot fi măsurate și cărora li se poate asocia un număr. Exemple de astfel de date pot fi înălțimea sau vârsta unei persoane, lungimea unui semifabricat, numărul de piese fabricate într-o zi etc.

Datele calitative pot fi observate dar nu pot fi măsurate direct și caracterizează subiectul lor din punct de vedere al calităților sale. Acestea pot fi descrise folosind atribute precum culoare, gen, textură, stare (bun / rău) etc. și se numesc *nominale*. Cu toate acestea unele pot avea o anumită ordine precum: mic, mediu, mare sau puțin important, important, foarte important, caz în care se numesc *ordinale*.

Pentru diferitele categorii de date se pot folosi anumite tipuri de date. Datele cantitative sunt reprezentate prin numere reale pentru mărimi sau numere întregi pentru numărări. În Excel valorile numerice pot fi clasificate ca fiind:

- Număr (Number)
- Monedă (Currency)
- Dată (Date)
- Timp (Time)
- Procent (Percentage)
- Frație (Fraction).

Datele calitative sunt de obicei reprezentate sub formă de text. În unele analize datelor ordinale li se pot atribui numere care să indice ordinea și rangul fiecărei variabile în raport cu celelalte. Aceste variabile nou obținute se numesc variabile fictive (*dummy variables*).



Atenție: atribuirea unor astfel de variabile fictive nu transformă datele calitative în date cantitative! Numerele au valoare simbolică

și nu pot fi folosite pentru calcule numerice precum media sau să fie folosite în operații aritmetice.

Exemplu: Atributele puțin important, important, foarte important pot fi codate folosind în ordine cifrele 1, 2 și 3.

În secțiunea următoare veți învăța să generați un șir de valori numerice aleatorii după o anumită distribuție. Conceptul de distribuție va fi clarificat în lucrări viitoare. Pentru moment este de ajuns să știți că distribuția este o lege care guvernează modul de aranjare a valorilor.

1.2.1. Generarea unui șir de numere aleatorii

În Excel 2010 generarea numerelor aleatorii se face cu ajutorul instrumentelor de analiză a datelor (*Data Analysis*) din panoul *Data*. Dacă acesta nu este prezent, adăugarea lui se face astfel:

- a. Selectați panoul *File* și apoi *Options*;
- b. Din fereastra de opțiuni, în partea stângă dați click pe *Add-Ins* iar în fereastra ce se deschide, în partea de jos, dați click pe butonul *Go...* având grijă ca în lista *Manage* din dreptul butonului să fie vizibil *Excel Add-Ins*;
- c. În fereastra *Add-Ins* bifați *Analysis ToolPak* (Pachet analiză de date), acesta fiind primul din listă, și dați click pe butonul *OK*;
- d. Dacă acesta este bifat și cu toate acestea *Data Analysis* nu este vizibil în panoul *Data*, debifați *Analysis ToolPak* din fereastra *Add-Ins* mai sus menționată, dați click pe butonul *OK* și repetați procedura începând cu punctul a.

Puteți vedea operațiunea de adăugare vizionând acest [video](#) în engleză.

Pentru a exemplifica folosirea generatorului de valori aleatorii, vom genera un șir de 20 de valori, așezate pe o coloană, normal distribuite, având media 8 și abaterea standard egală cu 0.1. Media și abaterea standard sunt elemente necesare generării după distribuția normală și vor fi explicate în detaliu în lucrarea 2.

Pașii sunt după cum urmează:

1. Selectați panoul *Data*, iar din secțiunea *Analysis* selectați *Data Analysis* (Instrumente de analiză a datelor);
2. Selectați din caseta de dialog care s-a deschis opțiunea *Random Number Generation* (Generare de Numere Aleatorii) și executați click pe butonul *OK*;
3. În fereastra de dialog care se deschide introduceți (Fig. 1.1.):
 - *Number of Variables* (Numărul coloanelor generate): **1**;
 - *Number of Random Number* (Numărul rândurilor generate): **20**;
 - *Distribution: Normal* (Distribuția, Normală);
 - *Mean* (Media): **8**;
 - *Standard Deviation* (Abaterea standard): **0.1**; aveți grijă la separatorul zecimal care poate fi punct sau virgulă în funcție de setările sistemului de operare. Dacă nu puteți trece la pasul următor (din cauza unei erori) este probabil din cauză că ați folosit separatorul zecimal greșit și trebuie să-l înlocuiți cu cel corect;
 - *Output Range* (Localizarea șirului generat), selectând/introducând celula de unde să înceapă localizarea șirului de date pe care urmează să-l generați;
4. Executați click pe butonul *OK*.

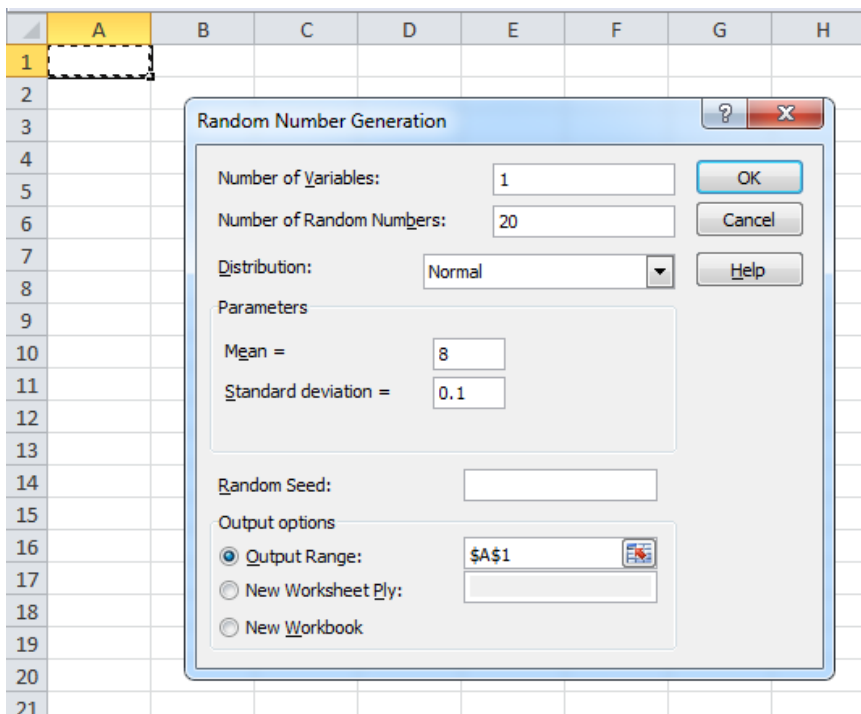


Fig. 1.1. Generarea numerelor aleatorii

În cazul selectării altor distribuții veți avea de ales alți parametri specifici distribuției selectate. Puteți afla mai multe informații despre generarea numerelor aleatorii din video-ul [acesta](#) în engleză sau citind [acest articol](#) în limba română, secțiunea *Random Number Generation*.

1.2.2. Noțiuni generale privind folosirea funcțiilor în Excel

Puterea programului Excel stă în capacitatea sa de procesare și manipulare a datelor. Acest lucru este deseori posibilă prin folosirea funcțiilor predefinite în program care au ca intrare anumiți parametri și pot executa calcule de la cele mai simple la cele mai complexe.

Introducerea funcțiilor în Excel se poate face în mai multe moduri. Două din cele mai comune sunt: prin scrierea directă a funcției într-o celulă sau prin folosirea Wizard-ului (program de asistență). Deoarece a doua modalitate oferă explicații pas cu pas pentru fiecare funcție în parte, în continuare vom detalia prima metodă.

La introducerea unei funcții într-o celulă, începem de fiecare dată cu semnul *egal* (=). Acest lucru îi semnalizează programului că urmează o funcție. După semnul *egal*, se introduc în ordine: numele funcției, o paranteză rotundă deschisă, lista de parametri necesari funcției separați de virgulă sau punct-virgulă (în funcție de setările de sistem) iar la final o paranteză rotundă închisă. Parametrii funcției pot fi alte funcții atâta timp cât rezultatul funcției folosite ca parametru respectă formatul parametrului pe care îl înlocuiește în funcția inițială. Sintaxa unei funcții arată similar cu:

$$= \text{funcție1}(\text{parametru1}, \quad \text{parametru2}, \\ \text{funcție2}(\text{param3}, \quad \text{param4}), \quad \dots)$$

Parametrii unei funcții sunt fie valori introduse manual fie o celulă sau un domeniu de celule selectate din foaia de lucru.

Exemplu: Pentru calculul mediei unui șir de numere folosim:

$$= \text{AVERAGE}(A2:A50)$$

Unde:

AVERAGE – este funcția folosită pentru calculul mediei aritmetice

A2:A50 – este domeniul de adrese a celulelor ce conțin valorile pentru care calculăm media respectiv de la celula A2 la A50.

Mai multe detalii privind folosirea de funcții și formule puteți obține vizionând [acest video](#) în limba română.

1.2.3. Sortarea valorilor

O funcție importantă este aceea de sortare a valorilor. Acest lucru se poate face prin accesarea funcției *Sort* (Sortare) din panoul *Data*. Înainte de apelarea funcției este recomandată selectarea datelor pe care doriți să le sortați. Se

poate selecta una sau mai multe coloane. Dacă selectați mai multe coloane, datele de pe celelalte coloane se rearanjează în funcție de coloana după care se face sortarea păstrându-se alinierea pe rânduri. Dacă spre exemplu, Tabelul 1.1. este sortat în funcție de nume, nu se modifică poziția relativă a vârstei și a orașului (Andrei va avea tot 14 ani și va fi tot din Cluj-Napoca)

Tab. 1.1. Exemplu de tabel pentru sortarea datelor

Nume	Vârstă (ani)	Oraș
Andrei	14	Cluj-Napoca
Maria	17	Zalău
Flaviu	15	Piatra Neamț
Ștefan	18	Oradea
Alexandru	13	Iași
Ana	12	Timișoara
Cristina	17	Brașov

Fereastra de sortare care se deschide (Fig. 1.2.) conține parametrii:

- Coloana după care se face sortare (Sort by);
- Criteriu de sortare (Sort on);
- Ordinea în care se face sortarea (Order).

Datele se pot selecta cu sau fără antetul tabelului (în exemplul de mai sus, rândul care conține: Nume, Vârstă (ani) și Oraș). În funcție de modul în care ați selectat datele, se bifează sau nu căsuța din dreapta sus *My data has headers* (Datele mele au antet).



Atenție: Dacă această opțiune rămâne bifată deși datele nu au antet, programul va considera primul rând de valori ca antet și va exclude primul rând din sortare.

Dacă opțiunea *My data has headers* nu este bifată, în lista de sortare după (Sort by) vor apărea identificatoarele coloanelor (A, B, C etc.) în locul antetului tabelului.

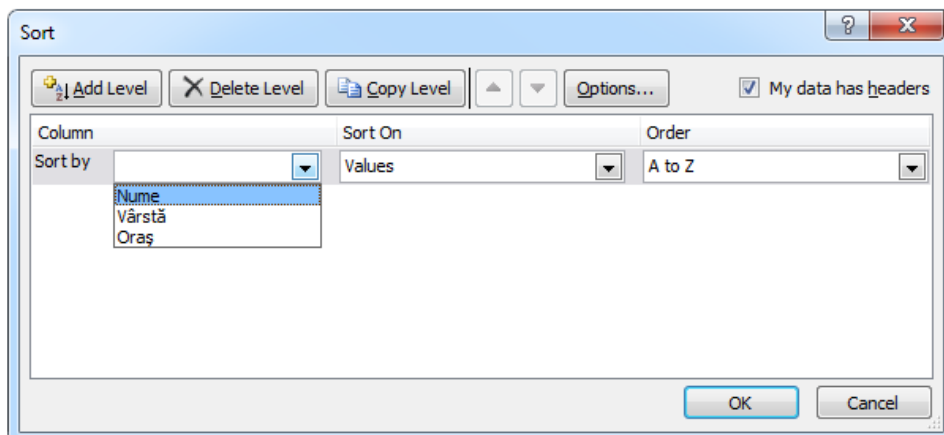


Fig. 1.2. Fereastra de sortare

Sortarea se poate face după valori, culoarea celulelor, culoarea textului sau a iconițelor prezente în celule. Ordinea poate fi crescătoare (A to Z), descrescătoare (Z to A) sau după o listă particularizată (Custom list). Mai multe informații despre sortarea datelor găsiți în engleză [aici](#) sau în română [aici](#).

1.2.4. Importarea datelor din surse externe

În Excel se pot importa date din diferite surse externe. Pentru a importa datele dați click pe panoul *Data* iar în secțiunea *Get external data* alegeți tipul corespunzător sursei datelor dvs. Pentru exemplificare vom folosi importul dintr-un fișier text.

Se dă click pe butonul *From Text* din secțiunea *Get external data* a panoului *Data*. Se deschide o fereastră care ne permite alegerea fișierului sursă. Odată selectat fișierul dorit, executăm click pe butonul *Import*, lansându-se astfel programul de asistență la import (Fig. 1.3.). Aici puteți selecta codificarea

fișierului dvs. (File origin - Originea fișierului), delimitatorul utilizat pentru a separa valorile dvs. (Delimiter - Delimitator) și metoda pe care Excel ar trebui să o utilizeze pentru a determina tipul de date (Data type Detection - Detectarea tipului de date). De cele mai multe ori valorile implicite funcționează bine și ar trebui să puteți face clic doar pe *Load* (Încărcare). În unele cazuri, poate fi necesar să modificați aceste opțiuni sau să vă editați datele și puteți face acest lucru aici, în funcție de caz dând click pe butonul *Edit*.

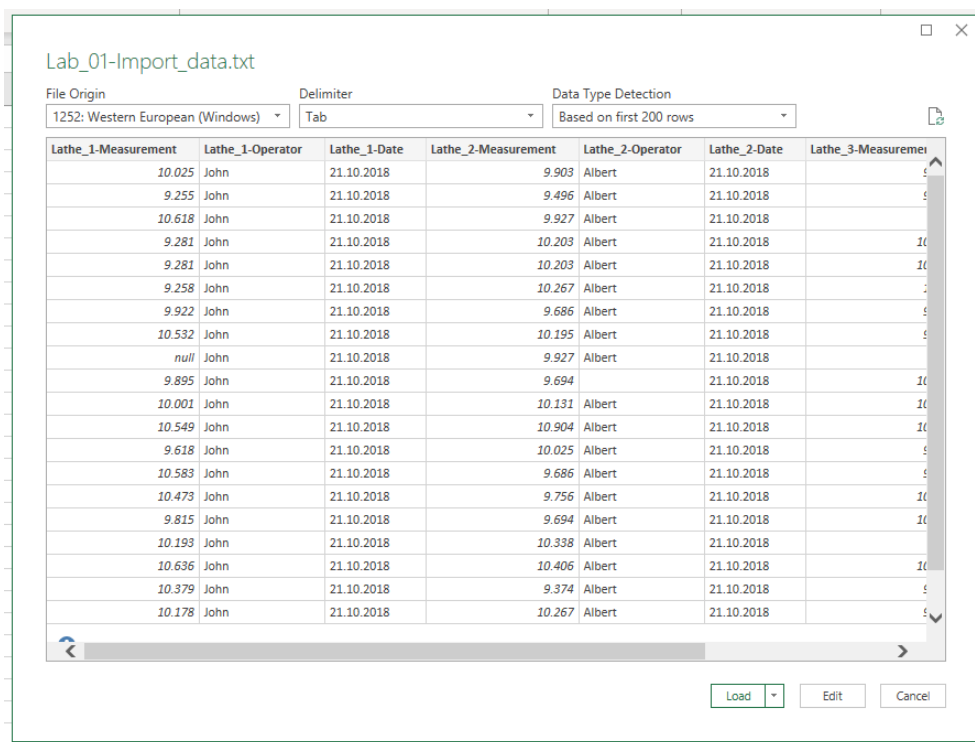


Fig. 1.3. Fereastra de import de date

Datele importate sunt plasate într-un format special de tabel, care are un rând de antet (Figura 1.4). Acest format consideră că datele sunt legate între ele de-a lungul tuturor coloanelor. De asemenea, conectează datele din fișierul Excel la fișierul sursă.

	A	B	C	D	E	F	G	H	I
1	Lathe_1-Measurement	Lathe_1-Operator	Lathe_1-Date	Lathe_2-Measurement	Lathe_2-Operator	Lathe_2-Date	Lathe_3-Measurement	Lathe_3-Operator	Lathe_3-Date
2	10.025	John	21.10.2018	9.903	Albert	21.10.2018	9.661	George	21.10.2018
3	9.255	John	21.10.2018	9.496	Albert	21.10.2018	9.889	George	21.10.2018
4	10.618	John	21.10.2018	9.927	Albert	21.10.2018	10	George	21.10.2018
5	9.281	John	21.10.2018	10.203	Albert	21.10.2018	10.128	George	21.10.2018
6	9.281	John	21.10.2018	10.203	Albert	21.10.2018	10.128	George	21.10.2018
7	9.258	John	21.10.2018	10.267	Albert	21.10.2018	10.35	George	21.10.2018
8	9.922	John	21.10.2018	9.686	Albert	21.10.2018	9.855	George	21.10.2018
9	10.532	John	21.10.2018	10.195	Albert	21.10.2018	9.625	George	21.10.2018
10		John	21.10.2018	9.927	Albert	21.10.2018	10	George	21.10.2018
11	9.895	John	21.10.2018	9.694		21.10.2018	10.102	George	21.10.2018
12	10.001	John	21.10.2018	10.131	Albert	21.10.2018	10.238	George	
13	10.549	John	21.10.2018	10.904	Albert	21.10.2018	10.066	George	21.10.2018
14	9.618	John	21.10.2018	10.025	Albert	21.10.2018	9.698	George	21.10.2018
15	10.583	John	21.10.2018	9.686	Albert	21.10.2018	9.711	George	21.10.2018
16	10.473	John	21.10.2018	9.756	Albert	21.10.2018	10.156	George	21.10.2018
17	9.815	John	21.10.2018	9.694	Albert	21.10.2018	10.705	George	21.10.2018
18	10.193	John	21.10.2018	10.338	Albert	21.10.2018	10.3	George	21.10.2018
19	10.636	John	21.10.2018	10.406	Albert	21.10.2018	10.684	George	21.10.2018
20	10.379	John	21.10.2018	9.374	Albert	21.10.2018	9.898	George	21.10.2018
21	10.178	John	21.10.2018	10.267	Albert	21.10.2018	9.795	George	21.10.2018
22	9.324	John	21.10.2018	9.661	Albert	21.10.2018	9.438	George	21.10.2018
23	11.391	John	21.10.2018	9.604	Albert	21.10.2018	9.873	George	21.10.2018
24	10.91	John	21.10.2018	9.467	Albert	21.10.2018	9.5	George	21.10.2018
25	10.845	John	21.10.2018	10	Albert	21.10.2018	10.667	George	21.10.2018
26	10.53	John	21.10.2018	10.174	Albert	21.10.2018	10.305	George	21.10.2018
27	10.464	John	21.10.2018	9.568	Albert	21.10.2018	9.726	George	21.10.2018
28	10.431	John	21.10.2018	9.833	Albert	21.10.2018	9.794	George	21.10.2018
29	9.403	John	21.10.2018	10.085	Albert	21.10.2018	10.098	George	21.10.2018

Fig. 1.4. Datele importate în formatul de tabel

Dacă doriți să schimbați tabelul într-un interval obișnuit de valori, trebuie să selectați mai întâi toate datele și în fila **Table tools** (vizibilă numai când datele din tabel sunt selectate) faceți clic pe *Convert to range* (Converțiți în interval).

1.2.5. Curățarea datelor

Pentru a extrage informații relevante și corecte din datele noastre, este nevoie să efectuăm o serie de operațiuni de „curățare” a datelor. Aceste operațiuni se referă în principal la eliminarea valorilor duplicate (în multe cazuri, dar nu neapărat în orice situație), rezolvarea problemei valorilor lipsă și a celor eronate.

În Excel, putem folosi funcția *Remove duplicates* din meniul *Data*, pentru eliminarea valorilor duplicat. Pentru acesta selectați întreg tabelul și dați click pe butonul *Remove duplicates* din meniul *Data*.

De multe ori cea mai simplă metodă de a rezolva problema valorilor lipsă sau a celor eronate este înlăturarea intrărilor care conțin valorile nule sau eronate. Dacă spre exemplu într-un experiment la o anumită oră nu s-a înregistrat corect valoarea temperaturii, acea intrare (cu valori măsurate, timp etc.) este eliminată.

Pentru a elimina rândurile cu valori lipsă puteți urma această procedură:

- a. Din panoul *Home* (Acasă), în secțiunea *Editing* (Editare), dați click pe *Find & Select* (Găsește și selectează) apoi pe *Go to special...* (Mergi la special). Din fereastra care se deschide (Fig. 1.5.) selectați *Blanks* (spații libere). Această operațiune va marca celulele goale din selecția dvs.

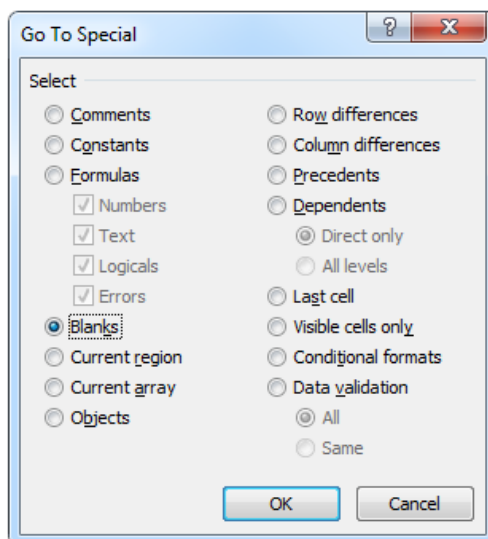


Fig. 1.5. Fereastra *Go To Special* cu *Blanks* selectat

- b. Odată selectate, rândurile care conțin aceste celule pot fi eliminate. Cu celulele goale încă selectate, dați click pe butonul *Delete* (Ștergere) din secțiunea *Cells* (Celule) a panoului *Home* apoi selectați comanda *Delete Sheet Rows* (Ștergerea rândurilor din foaie).

Mai multe informații despre eliminarea valorilor duplicat găsiți citind aceste articole în [română](#) sau [engleză](#) sau vizionând [acest video](#) în engleză.

1.2.6. Filtrarea datelor

Uneori avem nevoie doar de date care respectă anumite condiții sau au un anumit format. Pentru afișarea doar a datelor relevante de care avem nevoie, în Excel se pot folosi filtrele.

Pentru activarea filtrelor, se selectează datele pentru care se vrea aplicarea de filtre din panoul *Data*, secțiunea *Sort & Filter*, se dă click pe butonul *Filter*. La activarea filtrelor, în dreptul celulelor din antet (primul rând) apare câte o săgeată semnalizând prezența unei liste derulante. Dacă dați click pe săgeată, puteți vizualiza toate tipurile de filtre ce se pot aplica (Figura 1.6.). În funcție de formatul datelor pe care se aplică filtrele, opțiunile de filtrare pot diferi. Pentru valori numerice se aplică filtre numerice, pentru text cele de text ș.a.m.d.

Pentru valori numerice se pot stabili reguli de genul: egal cu ..., nu este egal cu ..., mai mic decât ..., mai mare decât ..., între două valori, mai mare sau mai mic decât media etc.

Pentru text se pot stabili reguli cum ar fi: egal cu ..., nu este egal cu ..., începe cu ..., se termină cu ..., conține ... etc.

Alternativ, din lista de valori se pot selecta doar valorile care se doresc a fi afișate, bifând valorile respective și debifându-le pe celelalte sau căutând anumite valori folosind căsuța de căutare (Search).

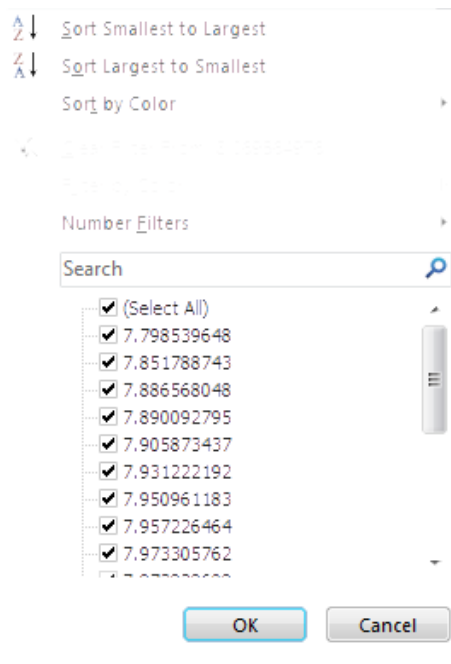


Fig. 1.6. Fereastra de filtre (Filter)

Mai multe informații despre filtre și acțiunea de filtrare puteți obține citind aceste articole în [română](#) sau [engleză](#) sau vizionând [acest video](#) în engleză.

1.2.7. Trasarea unei diagrame

În Excel există mai multe tipuri de grafice/diagrame. În această secțiune se va prezenta modul de trasare a unora dintre cele mai comune diagrame, respectiv: *Column* (coloane), *Line* (cu linii) *Bar* (diagramă cu bare) pentru un set de variabile.

Inserarea unei diagrame se face din panoul *Insert*, secțiunea *Charts* și conține următorii pași:

- a. *Selectarea datelor* care urmează a fi reprezentate; selectați doar valorile numerice relevante graficului. Toate cele trei tipuri de diagrame despre care discutăm au o singură variabilă reprezentată prin datele pe care vrem să le afișăm grafic. Aceste valori reprezintă înălțimea unei coloane (*Column*), distanța unui punct față de axa orizontală (*Line*) sau lungimea unei bare orizontale (*Bar*). Exemplu:

Dacă vrem să reprezentăm lungimea unor arbori, iar primul arbore are lungimea de 15 mm atunci prima coloană va avea înălțimea de 15 unități (*Column*), primul punct din diagrama liniară va fi la înălțimea de 15 unități (*Line*), iar bara orizontală va avea lungimea de 15 unități (*Bar*), așa cum este figurat în Fig. 1.7.

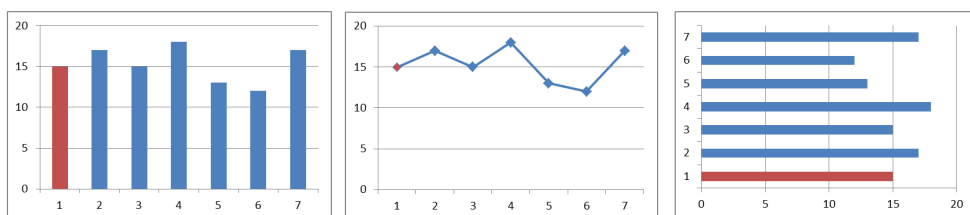


Fig. 1.7. Inserarea de grafice (*Column, Line, Bar*)

- b. *Inserarea tipului de grafic dorit.* Din panoul *Insert*, secțiunea *Charts* selectați tipul de grafic pe care vreți să-l inserați: *Column, Line* sau *Bar*. Obțineți un grafic similar unuia din Fig. 1.8.
- c. *Adnotarea diagramei.* Diagrama nu este completă fără etichetele fiecărei axe și titlul diagramei. Personalizarea graficului se face printr-un click pe *Chart Tools* din bara de titlu (dacă acesta nu este deja activ) iar din panoul *Design* se alege din secțiunea *Chart Layouts* (Disponerea elementelor graficului) un șablon care se potrivește cu ce avem nevoie. Pentru o personalizare mai avansată, din panoul *Layout* (Disponerea elementelor) se pot adăuga/particulariza fiecare element în parte. În final graficul arată similar celui din Fig. 1.9.

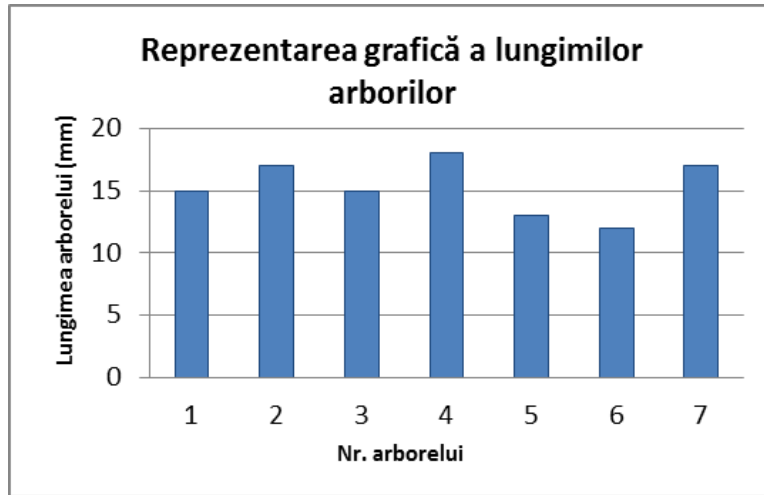


Fig. 1.8. Reprezentarea grafică după adăugarea tuturor elementelor

Mai multe informații despre inserarea de grafice puteți obține citind aceste articole în [română](#) sau [engleză](#) sau vizionând aceste video-uri în [română](#) sau [engleză](#).

1.3. Descrierea temei

Lucrați într-o companie producătoare de componente pentru industria auto, având funcția de specialist în calitate. După o revizie generală trei strunguri CNC reiau producția aceluiași tip de bucușă producând aprox. 500 de piese pe zi. Pentru acest tip de piesă, cota esențială este diametrul exterior al bucușei. Șeful Dvs. v-a cerut un raport detaliat privind producția pe cele trei strunguri și să determinați dacă revizia a avut ca rezultat o îmbunătățire a parametrilor de producției. Evaluarea producției va fi făcută prin prisma numărului de piese bune făcute înainte și după revizie pe fiecare strung, precum și dimensiunile maxime și minime ale pieselor obținute și amplitudinea.

Pentru realizarea raportului va trebui să faceți o eșantionare a producției de astăzi și să o comparați cu producția dintr-o zi de dinainte de revizie (date importate dintr-un fișier arhivă). În plus va trebui să curățați datele importate din arhivă pentru producția anterioară (eliminarea intrărilor dublate sau

având valori lipsă) și să găsiți ziua în care producția s-a desfășurat în condiții similare (aceeași operatori pe aceleași strunguri).

Pentru redactarea raportului trebuie:

- a. Să eșantionați, din producția de astăzi, câte 50 de piese pentru fiecare dintre cele trei strunguri și să măsurați diametrul exterior al fiecăreia.**

Vom simula eșantionarea pieselor prin generarea în Excel a unui șir de numere aleatorii așezat într-un tabel cu trei coloane (câte una pentru fiecare strung) și 50 de rânduri (pentru o valoare pentru fiecare piesă eșantionată). Tabelul va avea în total 150 de intrări. Datele vor fi normal distribuite având valoarea nominală (medie) 10 iar abaterea standard 0.XX unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15).

Pentru mai multe informații despre generarea unui șir de numere aleatorii vedeți secțiunea [Generarea unui șir de numere aleatorii](#).

- b. Pentru fiecare strung în parte determinați care este piesa cu diametrul exterior cel mai mic și care este piesa cu diametrul cel mai mare și amplitudinea.**

Pentru determinarea valorilor maxim și minim se pot folosi funcțiile $\min()$ și $\max()$ din Excel. Se vor determina în total șase valori (câte un maxim și un minim pentru fiecare din cele trei strunguri).

Amplitudinea se calculează scăzând din dimensiunea maximă pe cea minimă:

$$A = x_{max} - x_{min}$$

Detalii despre folosirea funcțiilor $\min()$ și $\max()$ în Excel găsiți în limba engleză [aici](#) sau pentru un sumar asupra folosirii funcțiilor în general, citiți secțiunea [1.2.2. Noțiuni generale privind folosirea funcțiilor în Excel](#).

c. Știind că diametrul exterior al bușei are toleranțele $10\pm 0.1\text{mm}$, câte piese bune s-au făcut pe fiecare mașină?

Pentru determinarea numărului de piese bune, având în vedere numărul mic de valori, o soluție este sortarea datelor pentru fiecare mașină în parte și numărarea valorilor care sunt în afara toleranțelor. Trebuie determinate trei valori, respectiv numărul de piese bune pentru fiecare din cele trei strunguri. Informații despre sortare găsiți în secțiunea [1.2.3. Sortarea valorilor](#).

d. Trebuie să comparați producția de astăzi cu producția dintr-o zi anterioară pentru a determina dacă revizia generală a îmbunătățit performanțele strungurilor. Pentru aceasta trebuie să importați din arhivă datele privind producția de dinainte de revizie.

Datele se află în fișierul text `Lucrarea 1 - Date_import.txt`. Dacă setările Dvs. de sistem folosesc ca separator zecimal virgula, deschideți fișierul `Lucrarea 1 - Date_import_virgula.txt`. Deschideți fișierul text într-un editor de text și analizați structura fișierului.

Importați datele într-o foaie nouă din cadrul aceleiași registru Excel folosind butonul *From text* (din fișier text) din cadrul panoului *Data*. În procesul de import se poate determina și formatul pentru fiecare coloană (vezi punctul e.)

Pentru mai multe informații despre cum se face importul dintr-un fișier text vedeți secțiunea [1.2.4. Importarea datelor din surse externe](#).

e. Odată inserate, datele trebuie formate corect și curățate.

Formatați datele folosind ca referință tabelul de mai jos:

Tab. 1.2. Formatele corespunzătoare fiecărei categorii de date

Strung 1, Strung 2, Strung 3	General
Operator 1, Operator 2, Operator 3	Text
Data 1, Data 2, Data 3	Data (short date)

Pentru a avea o imagine corectă asupra datelor noastre trebuie să rezolvăm problema valorilor lipsă. Cea mai simplă metodă este eliminarea intrărilor care au valori nule. În acest caz, pentru a păstra eşantioanele egale vom elimina întregul rând care conţine valori nule. O metodă simplă este identificarea valorilor și eliminarea manuală a fiecărui rând, această metodă fiind totuși mare consumatoare de timp. O metodă mai rapidă este descrisă în secțiunea [1.2.5. Curățarea datelor](#).

- f. **Știind că astăzi pe cele trei strunguri au lucrat operatorii din tabelul de mai jos, identificați zilele în care aceiași operatori au lucrat pe aceleași strunguri înainte de revizie și treceți-le într-un tabel similar celui de mai jos.**

Tab. 1.3. Operatorii pe cele trei strunguri

Strung	Operator	Data de azi	Data în care au lucrat aceiași operatori înainte de revizie
Strung 1	Georgescu		
Strung 2	Danescu		
Strung 3	Popescu		

Pentru identificarea datelor în care operatorii au lucrat în ordinea de mai sus, se pot folosi filtrele. Folosirea filtrelor este descrisă în secțiunea [1.2.6. Filtrarea datelor](#).

- g. **Sortați valorile obținute pentru data găsită și identificați câte valori se află în toleranțe pentru fiecare strung în parte (între 9.90 mm și 10.10 mm) și treceți-le în Raport.**

Sortarea se face în mod similar sortării de la pct. c. Trebuie să aveți trei valori, respectiv numărul de piese bune pentru fiecare din cele trei strunguri.

- h. **Comparați numărul de piese bune rezultate de pe fiecare strung înainte și după revizie. A avut un efect pozitiv revizia sau nu? Trasați**

o diagramă cu bare reprezentând numărul de piese bune din cele două date (înainte și după revizie) pentru fiecare strung și adnotați-o corespunzător.

Trasarea unei diagrame este expusă sumar în secțiunea [1.2.7. Trasarea unei diagrame](#). Pe diagramă trebuie să apară etichetele axelor și titlul diagramei.

i. Generați într-o foaie separată a aceluiași registru Excel, un raport similar cu cel din exemplu.

Formatarea și structura raportului din exemplu este doar pentru referință. Formatul raportului nu trebuie să fie identic cu cel din exemplu dar raportul trebuie să fie complet. Urmăriți cu atenție elementele pe care trebuie să le introduceți în raport. Nu toate datele cu care lucrați se vor introduce în raport (valorile măsurate, datele inserate din surse externe etc.). Raportul are rolul de sintetizare a informațiilor obținute.

Pentru a urmări mai ușor etapele de lucru puteți folosi lista de mai jos:

Sarcină	Unde trebuie să apară
a. Generarea unui șir de valori pe trei coloane cu 50 de intrări fiecare.	Foaie de lucru
b. Determinarea valorilor minime și maxime pentru fiecare din cele trei strunguri.	Raport
c. Determinarea numărului de piese bune din producția curentă pentru fiecare strung.	Raport
d. Importul de date din surse externe. Câte valori ați importat?	Foaie de lucru
e. Formatarea și curățarea datelor importate. Câte valori valide aveți după curățare (după eliminarea valorilor nule și dublurilor)?	Foaie de lucru
f. Identificarea zilerelor în care au lucrat aceeași operatori pe aceleași strunguri înainte de revizie.	Raport
g. Identificarea valorilor care se află în toleranțe pentru fiecare strung în parte pentru data dinainte de revizie.	Raport
h. Trasarea diagramei cu bare reprezentând numărul de piese bune din cele două date (înainte și după revizie) pentru fiecare strung și adnotați-o corespunzător.	Raport
i. Corectitudinea, aspectul și cât este de complet raportul	-

1.4. Resurse

Resursele menționate în această lucrare precum și alte resurse adiționale, le găsiți în tabelul de mai jos:

Descriere resursa	Limba	Format	Link
Generarea numerelor aleatorii	EN	video	https://youtu.be/Z_b3iSEbvtA?t=2m13s
Random number generation	RO	text	https://support.office.com/ro-RO/article/Analiza-datelor-7e71735c-c471-47e1-84ef-a8c23dc3098b
folosirea de funcții si formule	RO	video	https://www.youtube.com/watch?v=LYCGPdQ7mOE
Funcțiile min si max	EN	text	http://www.contextures.com/excelminmaxfunction.html
Sortarea datelor	EN	text	http://www.contextures.com/xlSort01.html
Sortarea datelor	RO	text	https://support.office.com/ro-ro/article/Sortarea-datelor-dintr-o-zon%C4%83-sau-dintr-un-tabel-62d0b95d-2a90-4610-a6ae-2e545c4a4654
Filtrarea datelor	RO	text	https://support.office.com/ro-ro/article/Filtrarea-datelor-dintr-o-zon%C4%83-sau-dintr-

			un-tabel-01832226-31b5-4568-8806-38c37dcc180e
Filtrarea datelor	EN	text	https://support.office.com/en-us/article/Filter-data-in-a-range-or-table-01832226-31b5-4568-8806-38c37dcc180e
Filtrarea datelor	EN	video	https://www.youtube.com/watch?v=zQMA9lw4EyU
Tipuri de date	RO	text	https://support.office.com/ro-RO/Article/Tipurile-de-date-din-modelele-de-date-e2388f62-6122-4e2b-bcad-053e3da9ba90
Tipuri de date	EN	text	https://support.office.com/en-us/article/Data-types-in-Data-Models-e2388f62-6122-4e2b-bcad-053e3da9ba90?rs=en-US&ui=en-US&ad=US
Validarea datelor	EN	text	http://www.excel-easy.com/basics/data-validation.html

Validarea datelor	EN	video	https://www.youtube.com/watch?v=uQm_CxwcOHw
Validarea datelor	RO	video	https://www.youtube.com/watch?v=hKjoILnjZEE
Validarea datelor	RO	text	http://www.drexcel.ro/articole/index.php/cum-pot-sa-validez-date-dupa-o-formula-in-microsoft-excel-2010/
Înlăturarea duplicatelor	EN	text	https://support.office.com/en-us/article/Filter-for-unique-values-or-remove-duplicate-values-d6549cf0-357a-4acf-9df5-ca507915b704#bremove_duplicate_values
Înlăturarea duplicatelor	RO	text	http://excelninja.ro/elimina-duplicate-in-excel/
Înlăturarea duplicatelor	EN	video	https://www.youtube.com/watch?v=vUbjWvGQTzg
Grafice	EN	text	http://www.excel-easy.com/data-analysis/charts.html
Grafice	EN	video	https://www.youtube.com/watch?v=-btUxQi76ql

Grafice	RO	video	https://www.youtube.com/watch?v=jNfEy79q-3U
Grafice	RO	text	http://excelninja.ro/grafice-in-excel-introductere/

2. Determinarea indicatorilor statistici de localizare și împrăștiere și reprezentarea grafică a datelor

- [Obiective de învățare](#)
- [Concepte teoretice](#)
- [Descrierea temei](#)
- [Resurse](#)

2.1. Obiective de învățare

Scopul acestei lucrări este familiarizarea studentului cu anumite concepte de bază de statistică descriptivă și a modului de folosire a acestora în practică. Pe lângă consolidarea abilității de generare a unui șir de numere aleatorii studentul va învăța să:

- Calculeze indicatori de localizare (medie, mediană, modală) și de împrăștiere (amplitudine, dispersie, abatere standard) atât folosind formule matematice cât și funcții din Excel;
- Împartă un șir de numere pe clase și să determine frecvența lor (absolută și relativă);
- Reprezinte grafic frecvențele prin trasarea unei histograme și a unei diagrame circulare;
- Genereze un raport în care sunt sintetizate informațiile culese și să tragă concluzii folosind rezultatele obținute.

2.2. Concepte teoretice

Pentru rezolvarea temei, aveți nevoie de câteva noțiuni de statistică descriptivă care vor fi prezentate sumar în această secțiune.

În general datele au două tendințe: de localizare și de împrăștiere. Localizarea ne arată în jurul cărei valori sunt grupate datele iar împrăștierea

ne indică cât de apropiate sau depărtate sunt valorile de valoarea în jurul căreia se grupează. Pentru cuantificarea acestor tendințe se folosesc o serie de indicatori statistici.

2.2.1. Indicatori de localizare

Un indicator de localizare este o valoare în jurul căreia tind să se grupeze valorile unui șir dat. De obicei acest indicator este reprezentat prin valoarea medie sau prin mediană și poate să aibă o valoare care aparține sau nu șirului. Există mai multe tipuri de medii fiecare fiind utilizată în cazuri specifice.

2.2.1.1. Media aritmetică

Media aritmetică este cel mai frecvent utilizat indicator de localizare. Se notează în general cu $M[x]$, iar în cazul mediei unei populație cu μ (miu) sau mediei unui eșantion cu \bar{x} .

Pentru un șir de valori x_1, x_2, \dots, x_n media aritmetică se calculează astfel:

$$M[x] = \frac{\sum_{i=1}^n x_i}{n} \quad (2.1.)$$

unde n este numărul de valori din șir.

2.2.1.2. Media geometrică

Media geometrică se notează cu M_g iar pentru un șir de valori x_1, x_2, \dots, x_n se calculează cu formula:

$$M_g = \sqrt[n]{x_1 * x_2 * \dots * x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad (2.2)$$

2.2.1.3. Media armonică

Media armonică se notează cu M_a iar pentru un șir de valori x_1, x_2, \dots, x_n se calculează astfel:

$$M_a = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (2.3)$$

2.2.1.4. Media pătratică

Media pătratică se notează cu M_p iar pentru un șir de valori x_1, x_2, \dots, x_n se calculează cu formula:

$$M_p = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} \quad (2.4)$$

Mediile aritmetică și pătratică sunt influențate de valorile mari ale șirului. Mediile geometrică și armonică sunt influențate de valorile mici și reduc din influența valorilor mari.

Între cei patru indicatori ai mediei există relația:

$$M_a < M_g < M[x] < M_p \quad (2.5)$$

2.2.1.5. Mediana

Mediana se notează cu M_e și este valoarea caracteristicii care ocupă locul central în șirul ordonat de valori, împărțindu-l în două grupe egale ca număr. Astfel, dacă șirul conține un număr *impar* de valori ($n=2k+1$) valoarea medianei este dată de a $(n+1)/2$ valoarea:

$$M_e = x_{\left(\frac{n+1}{2}\right)} \quad (2.6)$$

Spre exemplu, în *Tabelul 2.1.* mediana este valoarea 7.89 (marcată cu gri) deoarece șirul are 7 valori iar mediana împarte șirul în două grupe a câte trei valori fiecare (la stânga și la dreapta medianei).

Tab. 2.1. Mediana în cazul unui șir cu un număr impar de valori

7.80	7.85	7.89	7.89	7.91	7.93	7.96
------	------	------	------	------	------	------

În cazul unui număr par de valori mediana este egală cu media aritmetică a celor două valori centrale:

$$M_e = \frac{x_{(n/2)} + x_{(n/2+1)}}{2} \quad (2.7)$$

Dacă valorile sunt grupate pe clase, intervalul care conține elementul median se numește interval median sau clasă mediană.

În *Tabelul 2.2.* șirul are opt valori iar mediana se calculează prin realizarea mediei aritmetice a valorilor marcate cu gri (7.89 și 7.91). În acest caz mediana (egală cu 7.9) nu aparține șirului.

Tab. 2.2. Mediana în cazul unui șir cu un nr. par de valori

7.8	7.8	7.89	7.8	7.91	7.9	7.9	7.9
0	5		9		3	6	7

Mediana se determină mult mai ușor decât media și nu necesită nici un calcul. Din cauza ușurinței de determinare mediana este folosită frecvent în industrie la controlul statistic al fabricației. Mediana este preferată uneori mediei fiind mai puțin afectată de valorile extreme, fiind totodată mult mai stabilă la fluctuațiile de selecție decât media aritmetică.

2.2.1.6. *Modala*

Este valoarea caracteristicii corespunzătoare celei mai mari frecvențe. În cazul în care repartiția frecvențelor este reprezentată de o curbă, modala corespunde valorii maxime a caracteristicii. După cum repartiția experimentală are unul, două sau mai multe maxime, se numește unimodală, bimodală, sau multimodală (Fig. 2.1.). Dacă valorile sunt grupate pe clase, intervalul care conține elementul median se numește interval median sau clasă mediană. Pentru determinarea modalei este nevoie ca datele să fie împărțite pe clase, care este explicată în secțiuni viitoare.



Fig. 2.1. Distribuții unimodale, bimodale sau multimodale

2.2.1.7. Valoarea centrală

Valoarea centrală este media valorilor extreme. Aceasta se poate referi la un interval de grupare sau la întreg șirul. Se calculează astfel:

$$x_c = \frac{x_{\max} + x_{\min}}{2} \quad (2.8)$$

unde:

x_{\max} și x_{\min} sunt valorile extreme (maxim și minim) al șirului/intervalului pentru care se calculează valoarea central (x_c).

2.2.2. Indicatori de împrăștiere

Dacă indicatorii de localizare ne arată unde sunt datele noastre, indicatorii de împrăștiere ne arată cât de apropiate sau depărtate sunt datele față de valoarea în jurul căreia sunt grupate.

2.2.2.1. Dispersia

Unul dintre cei mai folosiți indicator ai variației (împrăștierii) datelor este dispersia. Acest indicator ține cont de toate valorile și de aceea caracterizează foarte bine întregul șir. În general este notată cu $D[x]$, dar dacă este vorba de dispersia unei populații, este notată cu σ^2 (sigma) iar în cazul unui eșantion cu s^2 . Formula de calcul a dispersiei unei populații cu valorile x_1, x_2, \dots, x_n este:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (2.9)$$

unde n este numărul elementelor din șir iar μ este media aritmetică a valorilor șirului.

Pentru un eșantion dispersia se corectează cu factorul $n/(n-1)$ astfel:

$$s^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2.10)$$

unde \bar{x} este media aritmetică a eșantionului.

2.2.2.2. Abaterea standard

Dacă scoatem rădăcina pătrată din dispersie obținem *abaterea standard* pentru populație:

$$\sqrt{\sigma^2} = \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (2.11)$$

și pentru eșantion:

$$\sqrt{s^2} = s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2.12)$$

Dacă se consideră abaterea față de altă constantă decât media aritmetică, atunci denumirea este de *abatere medie pătratică*.

2.2.2.3. Amplitudinea

Indică distanța dintre valoarea minimă și cea maximă a șirului:

$$R = x_{\max} - x_{\min}$$

2.2.3. Funcții Excel pentru calculul indicatorilor statistici

Pentru calculul indicatorilor statistici prezentați anterior se pot folosi în Excel funcții predefinite. Nu toți indicatorii se pot calcula cu o singură funcție, unii necesitând pași intermediari. Mai jos se vor prezenta ce funcții pot fi folosite și o scurtă descriere a modului lor de folosire. Pentru a obține rezultatele corecte, funcțiile trebuie să primească parametrii corecți.

Funcțiile Excel ce pot fi folosite sunt:

- **Media aritmetică:** =AVERAGE (A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm media aritmetică;
- **Media geometrică:** =GEOMEAN (A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm media geometrică;

- **Media armonică:** =HARMEAN(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm media armonică;
- **Media pătratică:** =SQRT(SUMSQ(A1:A10)/n), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm media pătratică, iar n este numărul de valori din șir. Funcțiile SQRT() și SUMSQ() sunt descrise în paragraful următor la **Rădăcina pătrată și Suma pătratelor**.
- **Mediana:** =MEDIAN(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm mediana;
- **Valoarea minimă și maximă:** =MIN(A1:A10) și respectiv =MAX(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm valorile minime și maxime;
- **Dispersia populației:** =VAR. P(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm dispersia;
- **Dispersia eșantionului:** =VAR. S(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm dispersia;
- **Abaterea standard a populației:** =STDEV. P(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm abaterea standard;
- **Abaterea standard a eșantionului:** =STDEV. S(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pentru care vrem să aflăm abaterea standard;

Următoarele funcții pot fi utile în calculul indicatorilor:

- **Produsul:** =PRODUCT(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pe care vrem să le înmulțim între ele;
- **Rădăcina pătrată (radicalul):** =SQRT(A1), unde A1 este adresa celulei care conține valoarea din care scoatem rădăcina pătrată (SQRT(4)=2);

- **Suma pătratelor:** =SUMSQ(A1:A10), unde A1:A10 sunt adresele celulelor care conțin valorile pe care vrem să le ridicăm pe fiecare la pătrat și apoi să le adunăm ($SUMSQ(1,2,3)=1+4+9=14$).
- **Ridicarea la putere:** =POWER(A1, n), unde A1 este adresa celulei care conține valoarea pe care o ridicăm la putere iar n este puterea la care ridicăm valoarea. Puterea poate fi exprimată și sub formă de fracție (ex. pentru radical de ordin n, valoarea puterii este $1/n$)

Unele calcule nu se pot face folosind funcții și trebuie făcute folosind operatorii clasici. Două exemple utile în această lucrare sunt:

- Diferența față de medie ($x_i - M_x$) în care trebuie determinată diferența fiecărei valori din șir față de medie;
- Calculul lui $1/x_i$, unde fiecare fracție trebuie calculată pentru fiecare valoare din șir.

O noțiune utilă este aceea de adresă absolută a unei celule. Adresa absolută a unei celule se folosește când într-o formulă de calcul se vrea folosirea aceleiași adrese ca argument chiar dacă formula este propagată în jos. Adresa absolută arată astfel: \$A\$1 (pentru celula A1) și se poate insera fie scriind semnul „\$” în fața literei și cifrei din adresa celulei fie având cursorul pe adresa dorită se apasă tasta F4. Informații suplimentare despre folosirea referințelor absolute și relative puteți afla citind articolul acesta în [engleză](#) sau în [română](#).

Pentru mai multe detalii legate de calculul indicatorilor puteți consulta următoarele resurse suplimentare:

- Indicatori de localizare: [Video RO](#) sau [Video EN](#)
- Dispersie și abatere standard:
 - Calculul dispersiei și abaterii standard [in Excel \(Video EN\)](#) sau [pe hârtie \(Video EN\)](#),
 - Funcțiile [stdev\(\)](#) și [var\(\)](#) ([Video EN](#))
 - Funcțiile [var.p\(\)](#) și [var.s\(\)](#) (text RO)
 - Funcțiile [stdev.p\(\)](#) și [stdev.s\(\)](#) (text RO)

2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvențe

Frecvența absolută este numărul de apariții al unui eveniment din cadrul unui experiment sau numărul elementelor cu aceeași valoare sau aparținând aceluiași interval.

Spre exemplu, dacă avem într-o ladă piese Lego® roșii, galbene și albastre, atunci frecvența pieselor galbene este dată de numărul de piese galbene din ladă, frecvența celor albastre de numărul celor albastre ș.a.m.d.

În cazul unor valori numerice, ca spre exemplu a unui șir de valori obținute prin măsurarea lungimii unor piese, este posibil să nu obținem de mai multe ori aceeași lungime. Atunci putem crea intervale de lungimi egale în care să putem grupa valorile măsurate.

Spre exemplu, dacă obținem valorile din Tabelul 2.3., putem să le împărțim în trei intervale respectiv:

- Între 6.50 și 7.50
- Între 7.50 și 8.50 și
- Între 8.50 și 9.50

Tab. 2.3. Valorile lungimilor pieselor măsurate (în mm)

6.97	7.01	7.46	8.01	8.50	8.91	9.12	9.43
------	------	------	------	------	------	------	------

În primul interval vom avea trei valori, în cel de-al doilea o valoare iar în ultimul patru valori acestea fiind frecvențele absolute ale celor trei intervale (tabel 2.4).

Tab. 2.4. Frecvențele valorilor din cele trei intervale

Intervalul	Frecvența absolută (a_i)
------------	------------------------------

[6:50, 7:50)	3
[7:50, 8:50)	1
[8:50, 9:50]	4

Trebuie avut grijă la valorile de capăt a intervalelor deoarece acestea pot fi valori care aparțin șirului pe care îl împărțim și nu vrem să le oitem sau să le numărăm de două ori (cum este cazul valorii 8.50). De la început trebuie stabilită regula de includere a acestora în intervale. Astfel dacă decidem ca la primul interval limita din dreapta să fie inclusă iar cea din stânga exclusă, aceeași regulă trebuie păstrată până la ultimul interval, excepție făcând valoarea maximă care trebuie inclusă în interval pentru a evita pierderea de valori.

Frecvența absolută se notează de obicei cu a_i unde i este numărului intervalului de grupare sau categorii (a_1 frecvența absolută a primului interval, a_2 a celui de-al doilea ș.a.m.d.). Dacă adunăm frecvențele absolute a fiecărui interval, obținem numărul total de elemente din șir.

Dacă vrem să vedem ce proporție reprezintă numărul de elemente dintr-un interval raportat la total, atunci calculăm **frecvența relativă**, cu formula:

$$f_i = \frac{a_i}{n} \quad (2.13)$$

unde f_i este frecvența relativă, a_i frecvența absolută iar n numărul de valori din șir.

În unele situații este nevoie să știm numărul de elemente până la o anumită valoare. Folosind frecvența absolută putem calcula **frecvența absolută cumulată** adunând frecvențele absolute a fiecărui interval până la valoarea stabilită. Frecvența cumulată poate fi crescătoare sau descrescătoare în funcție de sensul în care facem adunarea (începând cu primul interval sau respectiv cu ultimul).

Dacă din exemplul de mai sus vrem să vedem câte valori sunt strict mai mici de 8.50 (respectiv câte valori avem în primele două intervale) atunci calculăm **frecvența absolută cumulată crescătoare** adunând frecvențele primelor două intervale. Obținem valoarea 4 (3+1). Dacă vrem să obținem frecvența ultimelor două intervale (valori mai mari decât 7.50) atunci calculăm **frecvența absolută cumulată descrescătoare** și obținem valoarea 5 (4+1). Același lucru se poate face și pentru frecvența relativă, prin adunarea frecvențelor relative a fiecărui interval.

Tabelul de calcul al frecvențelor pentru șirul din exemplu arată ca *Tabelul 2.5*.

Tab. 2.5. Modul de calcul al frecvențelor pentru exemplul dat

Intervalul	Frecvența absolută (a_i)	Frecvența relativă (f_i)	Frecv. Abs. Cum. Cresc. (Ac_i)	Frecv. Rel. Cum. Cresc. (Fc_i)	Frecv. Abs. Cum. Dezcreșc. (Ad_i)	Frecv. Rel. Cum. Dezcreșc. (Fd_i)
[6:50, 7:50)	3	3/8	3	3/8	3+1+4	(3+1+4)/8
[7:50, 8:50)	1	1/8	3+1	(3+1)/8	4+1	(4+1)/8
[8:50, 9:50]	4	4/8	3+1+4	(3+1+4)/8	4	4/8

La un număr mare de valori, numărarea valorilor din fiecare interval devine foarte anevoioasă. De aceea se pot folosi în Excel anumite funcții de numărare:

- **Numărare simplă:** =COUNT (A1:A10) – numără câte valori numerice se află în celulele indicate (A1 până la A10);
- **Numărare condiționată (o singură condiție):** =COUNTIF (A1:A10, ">12") – numără din valorile selectate (A1 până la A10) câte dintre acestea respectă condiția ">12". Condiția poate fi enunțată în mai multe moduri. Câteva exemple includ:
 - "12" sau 12 – egal cu 12 (pot fi folosite și valori de tip text);

- *B1* – egale cu valoarea din celula *B1*;
- "*>12*" – strict mai mare decât *12*;
- "*>=12*" – mai mare sau egal cu *12*;
- "*<12*" – strict mai mare decât *12*;
- "*<=12*" – mai mare sau egal cu *12*;
- "*<>12*" – diferit de *12*.

Dacă se dorește să se facă referință la adresa unei celule ca valoare pentru condiție, atunci se poate scrie: "*>=&A1*" ceea ce semnifică „*valorile care sunt mai mari sau egale cu valoarea din celula A1*". Exemplu: =COUNTIF(A1:A10, "*>*"&B1) – numără din valorile aflate între celula *A1* și *A10* doar pe cele care sunt strict mai mari decât valoarea din celula *B1*.

- **Numărare condiționată (mai multe condiții simultan):**

=COUNTIFS(A1:A10, "*>12*", B1:B10, "*>*"&C1) - numără câte din valorile *A1* până la *A10* sunt mai mari ca *12* și în același timp câte valori din *B1* până la *B10* sunt strict mai mari decât valoarea din celula *C1*. Modul de folosire este similar funcției *COUNTIF*, doar că se pot folosi mai multe condiții îndeplinite simultan pentru șiruri multiple.

Un exemplu de calcul similar pentru determinarea frecvenței puteți găsi în acest [video](#) în limba română, cu observația că unele metode diferă de cele prezentate în această lucrare. Analizați cu atenție metodele și vedeți care dintre instrucțiuni se potrivesc la cerințele lucrării actuale.

Mai multe informații despre aceste funcții puteți găsi în acest video în [engleză](#) sau în [română](#).

2.2.5. Trasarea histogramei

Histograma este o reprezentare grafică a distribuției unui șir de date. În cazul nostru, pentru un număr de intervale, se determină numărul de elemente din fiecare interval și se reprezintă grafic sub forma unei diagrame cu bare verticale (*Fig. 2.2.*). Înălțimea fiecărei coloane este corespunzătoare

numărului de elemente din fiecare interval, iar lățimea este egală pentru toate deoarece intervalele sunt de lungimi egale.

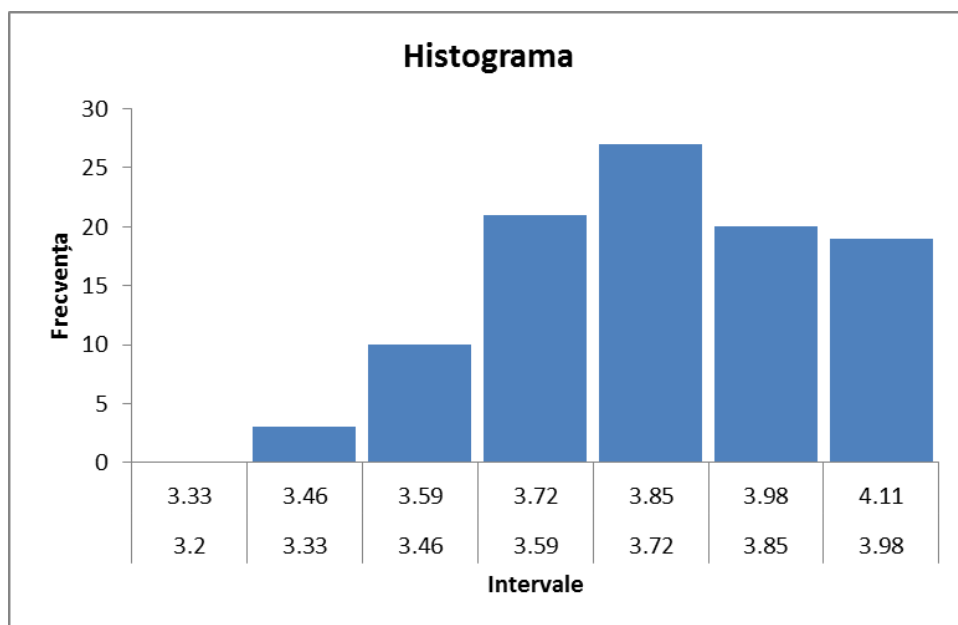


Fig. 2.2. Exemplu de histogramă

Dacă vrem să reprezentăm grafic frecvența absolută, atunci la inserarea unei diagrame cu bare verticale (*Column*) vom selecta din tabelul frecvențelor, datele care reprezintă frecvența absolută a fiecărui interval. Aceste valori vor da înălțimea coloanei corespunzătoare fiecărui interval.

Pentru situația în care nu avem intervale ci categorii (albastru, verde, roșu etc.) se procedează în mod similar.

În Excel histograma se poate genera automat cu ajutorul funcției *Histogram*, dacă sunt cunoscute șirul de date și capetele intervalelor de grupare. Din panoul *Data*, la *Data Analysis* găsim *Histogram*.

În fereastra *Histogram* (Fig. 2.3.) introducem parametrii:

- **Input range:** șirul de date pe care vrem să-l reprezentăm;

- **Bin range:** capetele intervalelor de grupare; tot ce este sub valoarea primul capăt și deasupra valorii celui de-al doilea vor fi introduse în categoriile *Less* (mai puține) și respectiv *More* (mai multe);
- **Output range:** locația unde vor fi poziționate rezultatele (tabelul de frecvențe și graficul);
- **Chart Output:** această opțiune trebuie bifată pentru crearea graficului.

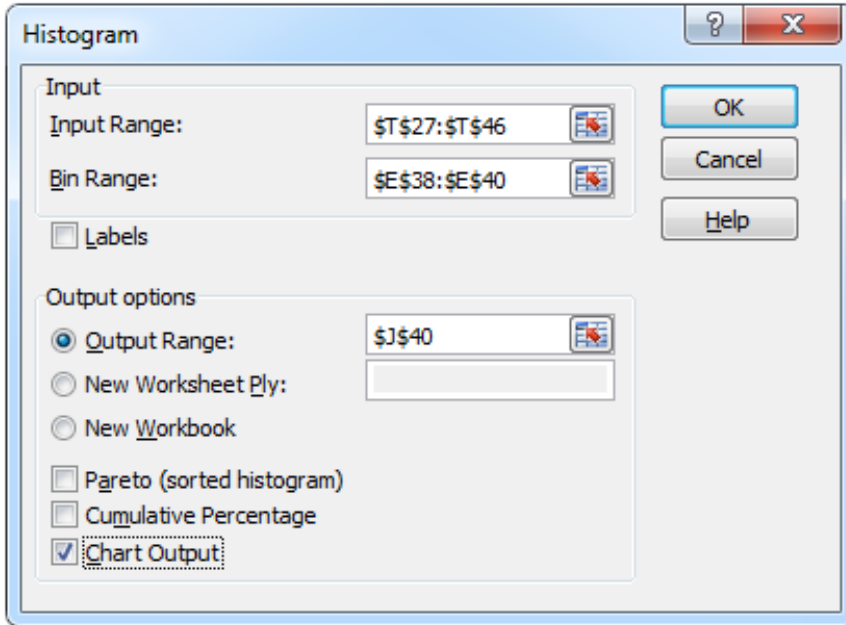


Fig. 2.3. Fereastra Histogram

Rezultatul obținut arată similar cu cel din Fig. 2.4.

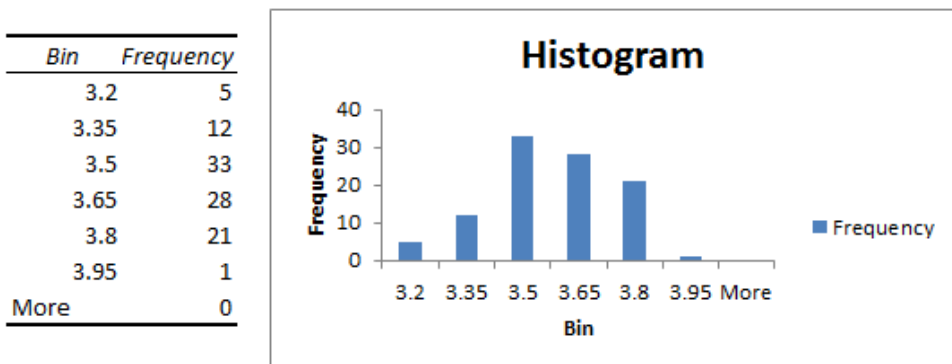


Fig. 2.4. Rezultatul obținut prin folosirea Histogram

Una din diferențele față de metoda „manuală” este faptul că limitele intervalelor apar o singură dată, acest lucru fiind totuși ușor de modificat.

2.3. Descrierea temei

Într-o secție de producție, există două linii de producție de rulmenți pentru un motor cu ardere internă. Cota esențială pentru montarea rulmentului este grosimea sa. Directorul de producție vă roagă să determinați performanțele celor două linii de producție și să întocmiți un raport în care să apară o analiză cantitativă a liniilor de producție. Pentru a descrie o linie de producție, trebuie calculați indicatorii statistici de localizare și de împrăștiere, creat un tabel de frecvențe și reprezentat grafic numărul de piese. Va trebui să determinați și câte dintre piesele produse sunt neconforme (prea subțiri sau prea groase).

Pentru aceasta trebuie să urmați pașii:

a. Generați două șiruri de date. Pentru a simula producția celor două linii trebuie să generați două șiruri de date astfel:

- Prima linie de producție are 100 de valori normal distribuite în jurul valorii medii de 3.65 mm având abaterea standard egală cu $0.1+0.XX$ unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15)
- A doua linie de producție are 100 de valori normal distribuite în jurul valorii medii $3.5+0.XX$ unde XX este ziua Dvs. de naștere iar abaterea standard egală cu 0.2.

Generarea se va face într-o foaie de calcul, într-un tabel similar cu *Tabelul 2.6.:*

Tab. 2.6. Exemplu de tabel pentru datele primare

Linia 1		Linia 2	
Piesa nr.	Grosimea (mm)	Piesa nr.	Grosimea (mm)
1 ... 100	Primul șir de numere generat	1 ... 100	Al doilea șir de numere generat

b. Determinați folosind formulele matematice specifice următorii indicatori pentru prima linie de producție și treceți-le în raport:

- Media aritmetică, geometrică, armonică și pătratică;
- Mediana ;
- Valoarea minimă și valoarea maximă;
- Valoarea centrală;
- Amplitudinea fiecărui șir;
- Dispersia populației și a eșantionului pentru fiecare șir;
- Abaterea standard populației și a eșantionului pentru fiecărui șir;

Formulele matematice ce trebuie aplicate pentru calculul indicatorilor de localizare pot fi revăzute în secțiunea [2.2.1. Indicatori de localizare](#), iar pentru indicatorii de împrăștiere în secțiunea [2.2.2. Indicatori de împrăștiere](#).

c. Faceți o verificare a calculelor făcute la punctul b. cu ajutorul funcțiilor existente în Excel.

Puteți folosi ca referință lista de funcții din secțiunea [2.2.3.Functii Excel pentru calculul indicatorilor statistici](#).

d. Împărțirea pe clase. Pentru aceste repere în firmă se face o împărțire în șapte (7) clase de calitate în funcție de grosime. Aceste clase se obțin conform Tabelului 2.7.

Tab. 2.7. Valorile limită pentru clasele de calitate

Clasă	Prea subțiri	Foarte subțiri	Subțiri	Medii	Groase	Foarte groase	Prea groase
Valori limită	[2,00 , 3.20]	(3.20 , 3.35]	(3.35 , 3.50]	(3.50 , 3.65]	(3.65 , 3.80]	(3.80 , 3.95]	(3.95 , 5,00]

Creați un tabel, similar celui din exemplu de raport, care să conțină valorile limită pentru cele șapte clase. Țineți cont de faptul că s-ar putea să aveți nevoie de valorile limită a fiecărui interval ca referință pentru alte funcții și drept urmare este recomandat să le scrieți în căsuțe individuale. Toate valorile mai mici de 3.2 intră în primul interval, iar cele mai mari de 3.95 în ultimul interval. Există posibilitatea să aveți și intervale fără nici o valoare.

- e. Determinarea frecvenței de apariție a pieselor în fiecare clasă pentru prima linie de producție. Fiecare dintre cele 100 de valori obținute se încadrează în una din clasele de la punctul d. Frecvența absolută reprezintă numărul de piese din fiecare clasă. Determinați frecvența absolută.**

Determinarea frecvenței absolute se poate face prin mai multe metode. Cea mai simplă este sortarea valorilor și numărarea valorilor care se încadrează între limitele fiecărei clase. Deoarece această metodă este mare consumatoare de timp, se recomandă folosirea funcțiilor din Excel. Pentru detalii legate de crearea tabelului de frecvențe vedeți secțiunea [2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvențe](#).

- f. Determinați frecvența absolută cumulată crescătoare și descrescătoare pentru prima linie de producție**

Determinarea frecvențelor cumulate se face conform instrucțiunilor din secțiunea [2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvențe](#).

- g. Determinarea frecvenței relative pentru rulmenții de pe prima linie de producție. Frecvența relativă reprezintă procentul de piese din fiecare clasă și apoi calculați frecvențele relative cumulate crescătoare și descrescătoare.**

Determinarea frecvențelor relative se face conform instrucțiunilor din secțiunea [2.2.4. Notiunea de frecvență și trasarea tabelului de frecvențe](#).

- h. Trasați o histogramă (folosind *Charts* » *Column*) care să afișeze câte piese se află în fiecare clasă din producția primei linii. În care clasă se află modala? Marcați această clasă pe grafic cu o altă culoare.**

Puteți să vă aduceți aminte despre cum se trasează grafice în Excel făcând referință la secțiunea [1.2.7. Trasarea unei diagrame](#).

Ce este o histogramă și care sunt particularitățile acesteia sunt descrise în secțiunea [2.2.5. Trasarea histogramei](#).

- i. Verificați corectitudinea histogramei de la punctul h. cu ajutorul funcției *Histogram*. Folosiți ca limite de intervale (*bins*) valorile claselor.**

Modul de folosire a funcției *Histogram* este descris în a doua parte a secțiunii [2.2.5. Trasarea histogramei](#).

- j. Trasați o diagramă circulară (*Pie*) care să figureze frecvența relativă pentru prima linie de producție și adnotați-o corespunzător.**

Puteți să vă aduceți aminte despre cum se trasează grafice în Excel făcând referință la secțiunea [1.2.7. Trasarea unei diagrame](#).

- k. Întocmiți un raport similar celui din exemplu în care să includeți elementele calculate mai sus (cele care trebuie introduse în raport). În secțiunea Rezultat a raportului scrieți concluziile Dvs. legate de linia de producție. Procesul este bun sau trebuie îmbunătățit? Ce recomandări puteți face având în vedere rezultatele obținute?**
- l. BONUS: Repetați operațiile de la punctele b. la j. pentru producția de pe linia a doua.**
- m. BONUS: Comparați cele două producții și determinați care dintre ele produce cele mai puține piese defecte, știind că înălțimea rulmenților trebuie să se încadreze între valorile 3.20 mm – 3.95 mm. Adăugați la raportul creat încă o pagină în care să centralizați (în mod similar**

primei linii de producție) rezultatele obținute pentru cea de-a doua linii de producție. Scrieți în secțiunea **Rezultat** o frază din care să reiasă care linie este mai eficientă (are mai puține piese rebut) și justificați cu date.

Pentru a urmări mai ușor etapele de lucru puteți folosi lista de mai jos:

Sarcină	Unde trebuie să apară
a. Generarea celor două șiruri de date.	Foaie de lucru
b. Determinarea următorii indicatori pentru fiecare dintre cele două linii de producție: <ul style="list-style-type: none"> • Mediile aritmetică, geometrică, armonică și pătratică (4 valori); • Mediana (1 valoare); • Valoarea minimă și valoarea maximă (2 valori); • Valoarea centrală (1 valoare); • Dispersia pentru fiecare șir (1 valoare); • Abaterea standard a fiecărui șir (1 valoare); • Amplitudinea fiecărui șir (1 valoare); 	Raport
c. Verificarea calculelor făcute la punctul b. cu ajutorul funcțiilor existente în Excel.	Foaie de lucru
d. Împărțirea pe clase (capetele fiecărui interval)	Raport
e. Determinarea frecvenței absolute pentru prima linie de producție	Raport
f. Determinarea frecvenței absolute crescătoare și descrescătoare	Raport
g. Determinarea frecvenței relative pentru prima linie de producție și a frecvențelor cumulate crescătoare și descrescătoare	Raport

h. Trasarea unei histograme care conține frecvența absolută (Bar Chart) și marcarea clasei care conține modala	Raport
i. Verificarea histogramei (Histogram)	Foaie de lucru
j. Trasarea unei diagrame circulare (Pie Chart)	Raport
k. Corectitudinea, aspectul și cât este de complet raportul	-
<i>l. Refacerea calculelor pentru linia a doua (BONUS)</i>	<i>Raport</i>
<i>m. Determinarea numărului de piese rebut și concluzionarea. (BONUS)</i>	<i>Raport</i>

2.4. Resurse

Descriere resursa	Limba	Format	Link
Exemplu de grupare a datelor pe intervale	RO	video	https://www.youtube.com/watch?v=q20oAPq6Djw
Adaugare Analysis Toolpak	EN	video	https://www.youtube.com/watch?v=2ujjkU-wAmc
Calculul frecventei	EN	video	https://www.youtube.com/watch?v=nh50bAKfj1o
Calculul frecventei	EN	video	https://www.youtube.com/watch?v=jbaXnT5CX18
Funcțiile min, max, sum, average, median, count, countif	EN	video	https://www.youtube.com/watch?v=9yPmlH-sRxQ
Funcții statistice	RO	video	https://www.youtube.com/watch?v=1MyAUs5Is8s
Abatere standard (esantion) funcția stdev.s	RO	text	https://support.office.com/ro-ro/article/STDEV-S-func%C8%9Bia-STDEV-S-7d69cf97-0c1f-4acf-be27-f3e83904cc23?ui=ro-RO&rs=ro-RO&ad=RO
Abatere standard (populatie) funcția stdev.p	RO	text	https://support.office.com/ro-ro/article/STDEV-P-func%C8%9Bia-STDEV-P-6e917c05-31a0-496f-ade7-4f4e7462f285?ui=ro-RO&rs=ro-RO&ad=RO

Dispersie (esantion) funcția var.s	RO	text	https://support.office.com/ro-ro/article/VAR-S-func%C8%9Bia-VAR-S-913633de-136b-449d-813e-65a00b2b990b?ui=ro-RO&rs=ro-RO&ad=RO
Dispersie (populatie) funcția var.p	RO	text	https://support.office.com/ro-ro/article/VAR-P-func%C8%9Bia-VAR-P-73d1285c-108c-4843-ba5d-a51f90656f3a?ui=ro-RO&rs=ro-RO&ad=RO
Abaterea standard și dispersia	EN	video	https://www.youtube.com/watch?v=efdRmGqCYBk
Abaterea standard și dispersia	EN	video	https://www.youtube.com/watch?v=gqOyy_NjflU
Funcțiile stdev și var	EN	video	https://www.youtube.com/watch?v=VzT9TgXN1ZE
Histogram (Data Analysis)	EN	video	https://www.youtube.com/watch?v=xe2Q6z6PTqk
Histogram (Data Analysis)	RO	text	https://support.office.com/ro-ro/article/Crearea-unei-histograme-b6814e9e-5860-4113-ba51-e3a1b9ee1bbe
Folosirea Descriptive statistics	EN	video	https://youtu.be/DrTTNwo-bjc?t=3m42s

Funcțiile count, countif, countifs	RO	video	http://www.lec-academy.ro/functii-de-numarare-partea-i/
Funcțiile count, countif, countifs	EN	video	https://www.youtube.com/watch?v=qIEgBHIPMg
Referințe absolute și relative	EN	text	http://www.gcflernfree.org/excel2013/15.2
Referințe absolute și relative	RO	text	http://www.competentedigitale.ro/excel/referinte.html

3. Eliminarea valorilor aberante și determinarea coeficientului de rebut

3.1. [Obiective de învățare](#)

3.2. [Concepte teoretice](#)

3.3. [Descrierea temei](#)

3.4. [Resurse](#)

3.1. Obiective de învățare

În această lucrare veți învăța despre eșantionarea aleatorie, cum se pot identifica și elimina valorile aberante și două moduri de determinare a coeficientului de rebut. În plus vă veți consolida cunoștințele despre frecvențe și crearea histogramelor.

3.2. Concepte teoretice

3.2.1. Eșantionarea

Eșantionarea este procesul de selecție a unităților statistice dintr-o populație statistică astfel încât, prin studierea eșantionului să putem formula concluzii legate de proprietățile populației din care s-a realizat eșantionarea.

De cele mai multe ori se preferă o *eșantionare aleatorie*. Unitățile sunt alese în mod aleatoriu din populație, adică toate elementele populației au o șansă egală de a fi incluse în eșantion. Chiar dacă eșantionul pe care îl luăm este aleatoriu, nu avem nici o garanție că acesta va caracteriza perfect populația din care eșantionăm. În unele cazuri, eșantioanele sunt foarte diferite de populație și de alte eșantioane, iar acest lucru se datorează *variației de eșantionare*.

În unele cazuri nu putem face eșantionări aleatorii. Spre exemplu, dacă avem de eșantionat dintr-un lot de blocuri de marmură de mari dimensiuni. În astfel de cazuri luăm un eșantion convenient încercând să evităm factori sistematici care ne-ar putea afecta eșantionul (în exemplul anterior nu vom lua doar blocuri care se află din colțurile lotului sau pe muchie, blocuri care sunt mai expuse riscului de lovire la manipularea lotului).

Pentru a face o eșantionare aleatorie în Excel, se folosește funcția *Sampling* (Eșantionare) din *Data Analysis* (aflat în panoul *Data*). Trebuie să introduceți (Fig. 3.1):

- Câmpul de celule care conține valorile sursă (*Input Range*) reprezentate prin cele 1000 de valori;
- Eșantionarea aleatorie (*Random sampling*);
- Numărul de unități eșantionate (*Number of samples*) respectiv 100;
- Adresa celulei unde vor fi poziționate cele 100 de valori eșantionate aleatoriu (*Output Range*).

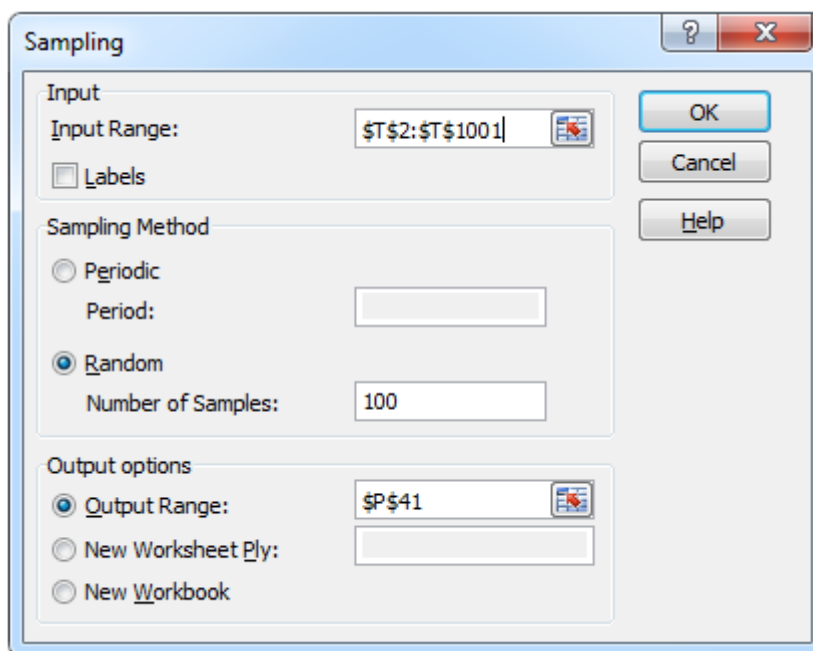


Fig. 3.1. Fereastra de eșantionare

3.2.2. Valorile aberante

Valorile aberante sunt valori care sunt foarte diferite față de celelalte valori ale șirului analizat. Aceste valori pot apărea fie din erori de înregistrare, corupere a datelor, probleme cu echipamentul de înregistrare etc. În astfel de cazuri este de dorit identificarea și eliminarea valorilor aberante, deoarece acestea pot modifica rezultatele obținute și să ne conducă să luăm decizii greșite.

Există mai multe metode de identificare și eliminare a valorilor aberante. Una dintre acestea este metoda cuartilelor. Quartilele sunt valorile care împart un șir ordonat de valori în sferturi (Fig. 3.1.). Există trei quartile: $Q1$, $Q2$ și $Q3$. Astfel între valoarea cea mai mică și prima cuartilă ($Q1$) se găsesc 25% din valori, între $Q1$ și $Q2$ se găsesc iar 25% din valori ș.a.m.d.

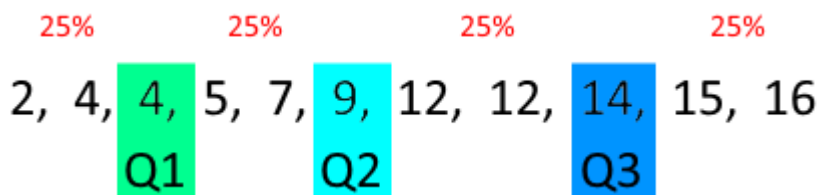


Fig. 3.2. Împărțirea unui șir în quartile

Pentru a folosi metoda cuartilelor este nevoie, în primă fază, să ordonăm crescător valorile șirului. Următorul pas este determinarea medianei care corespunde lui $Q2$. Ne aducem aminte că mediana este valoare care împarte șirul în două subșiruri de lungimi egale. În cazul exemplului din Fig. 3.2. mediana este valoarea 9. Dacă șirul are un număr par de valori, mediana este media aritmetică a celor două valori din mijlocul șirului.

În următoarea etapă determinăm valoarea lui $Q1$ și a lui $Q3$. Prima cuartilă se află la jumătatea subșirului de la stânga medianei iar a treia cuartilă la jumătatea subșirului de la dreapta medianei. Astfel obținem valoarea 4 pentru prima cuartilă și valoarea 14 pentru cea de-a treia.

Acum că avem determinate cele trei cuartile trebuie determinat intervalul intercuartilic (IQR – Interquartilic range). Intervalul intercuartilic este diferența între $Q3$ și $Q1$.

$$IQR = Q3 - Q1$$

În cazul nostru intervalul intercuartilic are valoarea:

$$IQR = 14 - 4 = 10$$

Următorul pas este determinarea limitelor interioare și a celor exterioare. Limitele interioare se determină prin înmulțirea intervalului intercuartilic cu 1.5 și adunarea acestei valori la $Q3$ și respectiv scăderea ei din $Q1$. Adică:

$$LI1 = Q1 - (1.5 * IQR)$$

$$LI2 = Q3 + (1.5 * IQR)$$

În cazul nostru $LI1=-11$ și $LI2=29$.

Calculul valorilor limită exterioare se face în mod similar, IQR fiind înmulțit cu 3 în loc de 1.5:

$$LE1 = Q1 - (3 * IQR)$$

$$LE2 = Q3 + (3 * IQR)$$

Valorile pentru șirul din exemplu sunt $LE1=-26$ și $LE2=44$.

Valorile aflate între $LI1$ și $LE1$ respectiv între $LI2$ și $LE2$ se numesc valori aberante minore. Cele care se află în afara valorilor limitelor exterioare se numesc valori aberante majore.

Aplicând această metodă putem determina dacă anumite valori sunt sau nu valori aberante. Valorile identificate ca fiind aberante pot fi eliminate dar nu este totdeauna necesar sau indicat să le eliminăm. Trebuie întâi să identificăm cauza valorilor aberante și abia apoi să luăm o decizie legată de eliminarea valorilor. Dacă valoarea aberantă apare ca urmare a unei erori de înregistrare spre exemplu, atunci se recomandă eliminarea acesteia, mai ales dacă influențează semnificativ parametrii eșantionului studiat (medie, abatere standard). Există totuși situații în care valorile aberante nu trebuie

eliminate. Un exemplu ar fi cazul unui experiment în care testăm influența compoziției unui material asupra rezistenței la tracțiune. Dacă o compoziție rezultă într-o valoare mult superioară a rezistenței la tracțiune, vrem să folosim compoziția respectivă în loc să o eliminăm. Un exemplu de folosire a metodei cuartilelor găsiți și în [acest articol în engleză](#). Mai multe despre cuartile puteți afla vizionând acest [video în engleză](#).

În Excel se poate folosi funcția =QUARTILE.INC() pentru determinarea cuartilelor. Argumentele funcției sunt adresele celulelor ce conțin șirul pentru care vrem să determinăm cuartila și a câta cuartilă (1, 2 sau 3).

Pentru mai multe detalii despre folosirea funcțiilor QUARTILE puteți citi acest [articol în engleză](#) sau [acesta în română](#).

3.2.3. Determinarea coeficientului de rebut

Coeficientul de rebut este procentul de piese neconforme rezultat în urma procesului de producție. Există mai multe metode de determinare a acestui coeficient, una din ele fiind cea grafică.

Pentru a determina grafic coeficientul de rebut trebuie să avem câteva informații la dispoziție:

- Tabelul frecvențelor;
- Histograma cu frecvența relativă a valorilor;
- Limitele de toleranță în afara cărora piesele sunt considerate rebut.

Tabelul frecvențelor este *Tab. 3.1*.

Tab. 3.1. Tabelul frecvențelor

<i>Intervale</i>		<i>ai</i>	<i>fi</i>
79.200	79.414	3	0.03
79.414	79.629	9	0.09
79.629	79.843	25	0.25
79.843	80.057	31	0.31
80.057	80.271	22	0.22
80.271	80.486	9	0.09
80.486	80.700	1	0.01

În primă fază reprezentăm grafic frecvența relativă prin trasarea histogramei. Pe histogramă marcăm poziția valorilor limită (toleranțele). Fiecare limită împarte aria histogramei în două regiuni: regiunea cu piese bune și cea cu piese defecte (regiunea de dinainte de limita inferioară și de după limita superioară reprezintă proporția de piese neconforme). Știind că aria totală a histogramei reprezintă 100% din piesele analizate, ariile din afara toleranțelor vor reprezenta procentul de piese neconforme (Figura 3.3.).

În exemplul din Fig. 3.4. aria cu albastru reprezintă piesele conforme iar ariile cu roșu reprezintă piesele neconforme. Observăm că limitele sunt $L_{inf}=79.3$ mm iar $L_{sup}=80.5$ mm. De pe axa verticală citim procentul de piese din fiecare interval (în primul interval avem 3% din piese iar în ultimul 1% din piese).

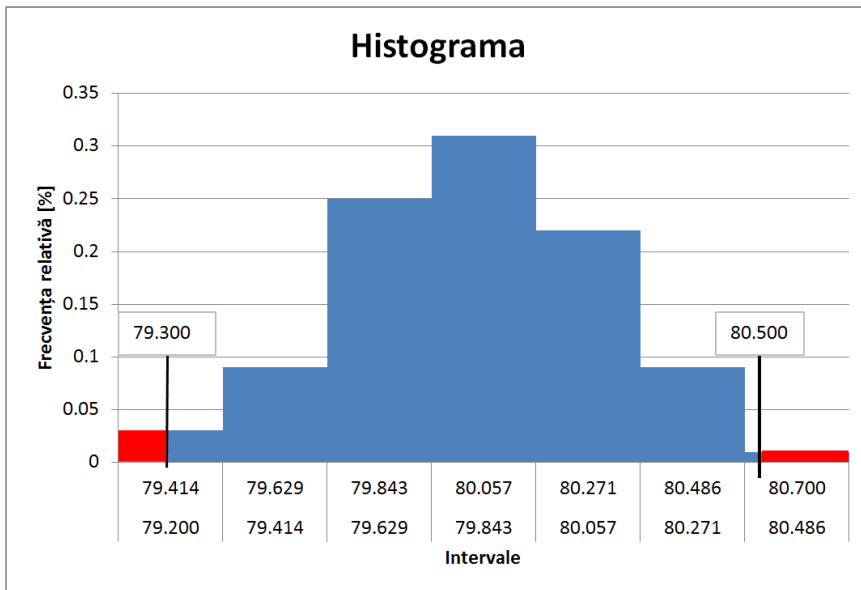


Fig. 3.3. Determinarea grafică a coeficientului de rebut

Pentru acest exemplu, coeficientul de rebut se determină astfel:

1. Determinăm cele două arii cu roșu:

- Știind că aria unui dreptunghi se determină prin înmulțirea lungimii celor două laturi, calculăm lungimea de pe axa orizontală scăzând din limita inferioară valoarea minimumului ($x_{inf}=79.3-79.2=0.1mm$). Dacă un interval are lungimea $d=0.214mm$, exprimăm lungimea calculată (x_{inf}) ca procent din lungimea (d) intervalului: $x_{inf\%} = 0.1/0.214=0.46$ respectiv 46%. Dacă înmulțim acest procent cu înălțimea coloanei din primul interval ($f_1=0.03$) obținem prima arie $A_{inf}=0.46*0.03= 0.014$ adică 1.4%. Acest ultim procent reprezintă proporția de piese mai mici decât limita inferioară.
 - În mod similar determinăm lungimea laturii orizontale a dreptunghiului din partea superioară ($x_{sup} =80.7-80.5=0.2mm$) și îl exprimăm procentual raportat la lungimea (d) a intervalului ($x_{sup\%} =0.2/0.214= 0.93$) sau 93%. Înmulțind cu frecvența relativă a ultimului interval ($f_7=0.01$) obținem cea de-a doua arie $A_{sup}=0.93*0.01= 0.009$ adică 0.9%.
2. Coeficientul de rebut este suma celor două arii ($A_{inf} + A_{sup}=0.014+0.009=0.023$). Coeficientul de rebut este deci 2.3%.

3.3. Descrierea temei

Sunteți rugat de directorul de producție să faceți o analiză a producției de carcase turnate din luna ianuarie 2020. Cota urmărită este înălțimea carcusei. Dintr-o arhivă trebuie să extrageți valorile reprezentând înălțimea carcusei. Din cele 1000 de valori veți eșantiona 100 și veți face calcule statistice pe acest eșantion. Pentru ședința ce va avea loc săptămâna viitoare trebuie să întocmiți un raport care conține statisticile ce descriu producția (medie, abatere standard etc.) tabelul frecvențelor, histograma și coeficientul de rebut pentru luna ianuarie. Înainte de a prezenta datele, trebuie identificate și eliminate valorile aberante ale eșantionului analizat.

Aveți de urmat acești pași:

- a. **Importați din fișierul *Lucrarea 3 - Date_import.txt* coloana reprezentând înălțimea carcusei. Dacă separatorul zecimal este virgula folosiți fișierul *Lucrarea 3 - Date_import_virgula.txt***

Importul datelor din fișierul text este explicat în *Lucrarea 1* secțiunea [1.2.4. Importarea datelor din surse externe](#).

- b. **Din populația totală de 1000 de piese eșantionați aleatoriu 100 de valori și introduceți-le într-o foaie de lucru.**

Eșantionarea se face conform instrucțiunilor din secțiunea [3.2.1. Eșantionarea](#).

- c. **Determinați quartilele (Q1, Q2 și Q3) și apoi calculați intervalul intercuartilic (IQR).**

Determinarea quartilelor este explicată în secțiunea [3.2.2. Valorile aberante](#).

- d. **Vedeți dacă aveți valori aberante (prin determinarea limitelor LI) iar dacă aveți, eliminați-le.**

Identificarea valorilor aberante cu metoda quartilelor este prezentată în a doua parte a secțiunii [3.2.2. Valorile aberante](#). Pentru acest exercițiu trebuie

să determinați doar LI1 și LI2 și să eliminați valorile dinafara acestora (mai mici decât LI1 și mai mari decât LI2).

- e. După eliminarea valorilor aberante identificați valoarea minimă și cea maximă și calculați amplitudinea, media aritmetică, dispersia și abaterea standard a eșantionului, folosind funcțiile corespunzătoare din Excel.**

Determinarea indicatorilor statistici se face prin aplicarea metodelor prezentate în [Lucrarea 2. Determinarea indicatorilor statistici de localizare și împrăștiere și reprezentarea grafică a datelor](#)

- f. Pentru șirul fără valori aberante, creați un număr de 7 intervale egale cuprinse între valoarea minimă și valoarea maximă a eșantionului.**
- g. Determinați frecvența absolută și cea relativă pentru fiecare interval creat și întocmiți tabelul frecvențelor.**

Determinarea frecvențelor a fost explicată în Lucrarea 2, secțiunea [2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvente](#)

- h. Trasați o histogramă care să conțină frecvența relativă.**

Trasarea histogramei a fost explicată în Lucrarea 2, secțiunea [2.2.5. Trasarea histogramei](#)

- i. Știind că înălțimea carcasei trebuie să se încadreze între valorile de 78mm și 82mm, determinați coeficientul de rebut (procentul de piese neconforme)**

Determinarea coeficientului de rebut este explicată în [3.2.3. Determinarea coeficientului de rebut.](#)

- j. BONUS: Curățați întreaga populație de piese (cele 1000) și apoi pentru întreaga populație:**
- calculați indicatorii statistici de la punctul c. (*Atenție: se vor folosi funcțiile pentru populație în locul celor pentru eșantion*);
 - faceți tabelul frecvențelor;
 - reprezentați grafic frecvența relativă.

- k. BONUS: Există vreo diferență între indicatorii obținuți pentru eșantion și cei pentru populație? Putem aproxima distribuția populației cu cea a eșantionului? Notați aceste observații în raport.
- l. BONUS: Extrageți 3 eșantioane de 100 de valori folosind metoda de eșantionare folosită anterior. Trasați o diagramă boxplot similară cu cea de mai jos și interpretați-o.

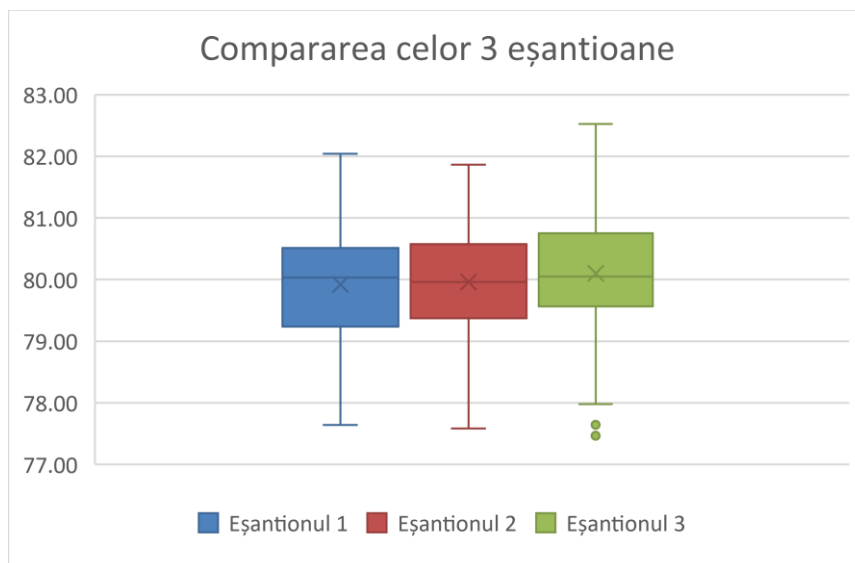


Fig. 3.4. Exemplu de diagramă Boxplot

Pentru a urmări mai ușor etapele de lucru puteți folosi lista de mai jos:

Sarcină	Unde trebuie să apară
a. Importați coloana reprezentând diametrul interior al cilindrilor.	Foaie de lucru
b. Din populația totală de 1000 de piese eșantionați aleatoriu 100 de valori și introduceți-le într-o foaie de lucru.	Foaie de lucru
c. Determinați cuartilele (Q1, Q2 și Q3) și apoi calculați intervalul intercuartilic (IQR).	Raport
d. Determinați dacă aveți valori aberante iar dacă aveți, eliminați-le.	Foaie de lucru
e. Pentru șirul fără valori aberante determinați valoarea minimă și cea maximă și calculați amplitudinea, media aritmetică, dispersia și abaterea standard a eșantionului, folosind funcțiile corespunzătoare din Excel.	Raport
f. Creați un număr de 7 intervale egale între valoarea minimă și valoarea maximă a eșantionului.	Raport
g. Determinați frecvența absolută și cea relativă pentru fiecare clasă creată și întocmiți tabelul frecvențelor.	Raport
h. Trasați o histogramă care să conțină frecvența absolută.	Raport
i. Știind că toleranța pentru diametru interior este $80 \pm 0,5\text{mm}$, determinați coeficientul de rebut (procentul de piese neconforme)	Raport
j. BONUS: Curățați întreaga populație de piese (cele 1000) și apoi pentru întreaga populație: <ul style="list-style-type: none"> • calculați indicatorii statistici de la punctul c.; • faceți tabelul frecvențelor; • reprezentați grafic frecvența absolută. 	Raport
k. BONUS: Există vreo diferență între indicatorii obținuți pentru eșantion și cei pentru populație? Putem aproxima	Raport

distribuția populației cu distribuția eșantionului? Notați aceste observații în raport.	
I. BONUS: Extrageți 3 eșantioane de 100 de valori folosind metoda de eșantionare folosită anterior. Trasați o diagramă boxplot similară cu cea de mai jos și interpretați-o.	Raport

3.4. Resurse

Descriere resursa	Limba	Format	Link
Eșantionarea în Excel	EN	video	https://www.youtube.com/watch?v=5XrJcFmbpWI
Determinarea valorilor aberante	EN	text	http://www.wikihow.com/Calculate-Outliers
Cuartile, valori aberante, boxplot	EN	video	https://www.youtube.com/watch?v=gkI-HeMiKzQ
Metoda cuartilelor	EN	text	http://www.wikihow.com/Calculate-Outliers
Funcția QUARTILE.INC	EN	text	https://support.office.com/en-us/article/QUARTILE-INC-function-1bbacc80-5075-42f1-aed6-47d735c4819d
Funcția QUARTILE.INC	RO	text	https://support.office.com/ro-RO/article/QUARTILE-INC-func%C8%9Bia-QUARTILE-INC-1bbacc80-5075-42f1-aed6-47d735c4819d
Eșantionarea în Excel	EN	video	https://www.youtube.com/watch?v=5XrJcFmbpWI
Determinarea valorilor aberante	EN	text	http://www.wikihow.com/Calculate-Outliers

Cuartile, valori aberante, boxplot	EN	video	https://www.youtube.com/watch?v=qkl-HeMiKzQ
Metoda cuartilelor	EN	text	http://www.wikihow.com/Calculate-Outliers
Funcția QUARTILE.INC	EN	text	https://support.office.com/en-us/article/QUARTILE-INC-function-1bbacc80-5075-42f1-aed6-47d735c4819d
Funcția QUARTILE.INC	RO	text	https://support.office.com/ro-RO/article/QUARTILE-INC-func%C8%9Bia-QUARTILE-INC-1bbacc80-5075-42f1-aed6-47d735c4819d

4. Estimarea parametrilor populației

4.1. [Obiective de învățare](#)

4.2. [Concepte teoretice](#)

4.3. [Descrierea temei](#)

4.4. [Resurse](#)

4.1. Obiective de învățare

În această lucrare veți învăța cum se estimează parametrii unor populații cu diferite distribuții (media și abaterea standard) și cum se interpretează rezultatele obținute. Veți învăța să folosiți tabelele de repartiție și să înțelegeți interpretarea acestora.

4.2. Concepte teoretice

În multe situații nu putem măsura totalitatea elementelor unei populații statistice. De aceea se recurge la o metodă numită *eșantionare*, care constă în prelevarea unui grup de elemente din populația studiată. Prelevarea grupului se face de obicei aleatoriu, iar grupul se numește *eșantion*.

Analizând eșantionul prelevat putem trage concluzii legate de parametrii necunoscuți ai populației analizate. Cu cât eșantionul este mai mare, sau cu cât luăm mai multe eșantioane, cu atât putem estima parametrii populației mai bine. Analiza eșantionului o facem prin prisma unor funcții caracteristice variabilelor aleatorii din eșantion, funcții numite *statistici*. Plecând de la aceste statistici putem estima parametrii populației. Distribuția probabilității unei statistici se numește distribuția de eșantionare.

Distribuția de eșantionare depinde de distribuția populației, mărimea eșantionului și de metoda de eșantionare.

În cazul în care caracteristicile eșantionului sunt aproximativ egale cu cele ale populației, este determinat un eșantion reprezentativ.

Pentru efectuarea estimăției, statistica, regula sau formula aleasă poartă denumirea de *estimator*. Estimatorul poate fi o variabilă aleatorie dependentă de eșantion.

Estimarea poate fi:

- *punctuală* – dacă parametrul ϵ al unei populații se estimează printr-o valoare izolată determinată de un estimator e . Estimația punctuală nu coincide cu adevărata valoare a parametrului dar este foarte apropiată. Parametrii populației pot fi estimați cu ajutorul mai multor estimatori diferiți și în consecință poate avea precizii diferite.
- *cu interval de încredere* – dacă se stabilește un interval $(e-\Delta, e+\Delta)$ care cu o probabilitate dată P , să includă valoarea adevărată a parametrului estimat ϵ .

Estimarea punctuală are ca dezavantaj precizia necorespunzătoare datorită necunoașterii domeniului de variație a parametrului estimat determinat de probabilitatea estimării. În industrie domeniul de variație a unei caracteristici reflectă și nivelul de calitate al fabricației.

Probabilitatea este dependentă de intervalul de variație stabilit prin apartenența unei variabile aleatorii la un interval dat, dată de relația:

$$P = P(a \leq X \leq b) = \int_a^b f(x) dx$$

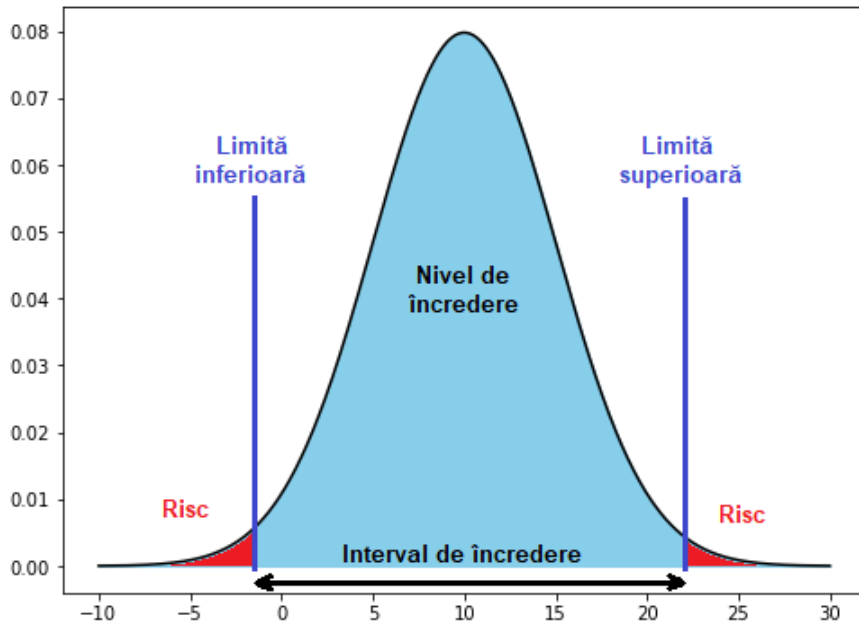


Fig. 4.1 Legătura între intervalul de încredere, nivel de încredere și risc

S-a stabilit astfel un interval (a, b) , *interval de încredere* care are proprietatea de a conține adevărata valoare a parametrului, cu o probabilitate P (Figura 4.1). Notând cu x_0 adevărata valoare a parametrului pentru care este estimată valoarea x , abaterea $(x_0 - x)$ este mai mică decât o valoare ε cu o probabilitate β foarte mare (90% - 95%).

$$P(|x - x_0| < \varepsilon) = \beta$$

$$P(x - \varepsilon < x_0 < x + \varepsilon) = \beta$$

Valoarea ε poate fi determinată pentru o anumită probabilitate β astfel încât valoarea adevărată x_0 să aparțină intervalului $(x - \varepsilon, x + \varepsilon)$ cu probabilitatea β . Deci se poate construi un interval (a, b) în care $a = x - \varepsilon$, și $b = x + \varepsilon$, denumit interval de încredere. Intervalul complementar se numește "*interval de semnificație*", (Figura 4.1). Intervalul de încredere poate fi:

- bilateral simetric, utilizat în tehnica măsurătorilor unde abaterile mărimilor în sensul pozitiv și negativ sunt egale ($A \pm T$), (Figura 4.1);

- bilateral asimetric, utilizat în tehnica măsurătorilor unde abaterile mărimilor în sensul pozitiv și negativ nu sunt egale;
- unilateral are riscul plasat în stânga sau dreapta, definit de una din relațiile: $P(x_0 < a) = \alpha$ (risc unilateral stânga), $P(x_0 > a) = \alpha$ (risc unilateral dreapta). Se utilizează în tehnologia montajului cu joc unde arborele are toleranța negativă în timp ce alezajul are toleranța pozitivă, sau la elaborarea planurilor de control statistic unde numărul defectelor trebuie să nu depășească o anumită limită superioară.

4.2.1. Estimarea mediei

4.2.1.1. Estimarea mediei când dispersia este cunoscută

Dintr-un lot a cărui dispersie σ^2 este cunoscută, (cazul furnizorului), se extrage un eșantion de volum n pentru care se calculează media \bar{x} . Cu riscul α se estimează media lotului μ .

Statistica utilizată este:

$$z = \frac{\mu - \bar{x}}{\sigma/\sqrt{n}}$$

care conduce la o repartiție normală .

Valorile z se citesc din tabelul distribuției normale (Anexa 2)

Risc bilateral simetric, RBS

Se cunosc: $n, \bar{x}, \sigma^2, \alpha$.

Se determină: μ .

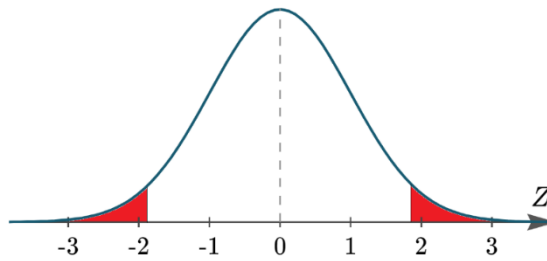


Fig. 4.2. Distribuția normală cu Risc Bilateral Simetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1 - \alpha$$

$$P = P\left(-z_{\alpha/2} \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

Din tabelul repartiției normale standard se aleg valorile $\pm z_{\alpha/2}$. Riscul este de ambele părți și este egal cu jumătate din α (Figura 4.2). Din inegalitate se estimează intervalul în care se află media lotului:

$$\mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 16$;
- $\bar{x} = 10,1$;
- $\sigma^2 = 0,01$;
- $\alpha = 10\%$ (RBS)

Se determină: $\mu = ?$

Pași de lucru:

$$P = P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 90\%$$

$$P = P\left(-z_{\alpha/2} \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 90\%$$

Din tabelul repartiției normale $N(0,1) \Rightarrow \pm z_{\alpha/2} = 1,645$

$$\mu = 10,1 \pm 1,645 \frac{0,1}{\sqrt{16}}$$

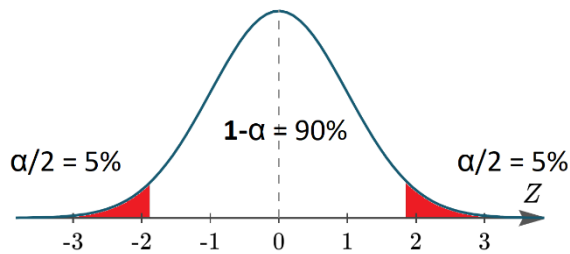


Fig. 4.3. Exemplu de calcul cu distribuția normală și RBS

Risc bilateral asimetric, RBA

Se cunosc: $n, \bar{x}, \sigma^2, \alpha$.

Se determină: μ .

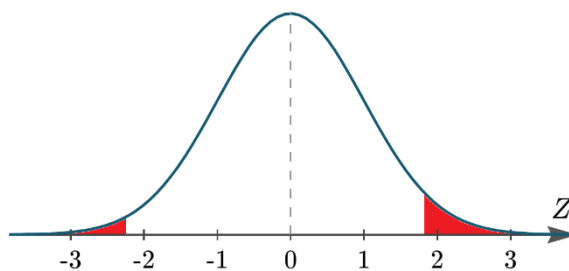


Fig. 4.4. Distribuția normală cu Risc Bilateral Asimetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-z_s \leq z \leq z_d) = 1 - \alpha$$

$$P = P\left(-z_s \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_d\right) = 1 - \alpha$$

Din tabelul repartiției normale standard se aleg valorile z_s ; z_d . . Riscul este de inegal și avem drept urmare un risc la stânga diferit de cel de la dreapta (Figura 4.4). Din inegalitate se estimează intervalul în care se află media lotului:

$$\bar{x} - z_s \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_d \frac{\sigma}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 20$;
- $\bar{x} = 14,95$;
- $\sigma^2 = 0,01$
- $\alpha_s = 3\%$; $\alpha_d = 6\%$ (RBA)

Se determină: $\mu = ?$

Pași de lucru:

$$P = P(-z_s \leq z \leq z_d) = 91\%$$

$$P = P\left(-z_s \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_d\right) = 91\%$$

Din tabelul repartiției normale $N(0,1) \Rightarrow z_s = -1.880$, $z_d = 1.555$

$$14,95 - 1,880 \frac{0,1}{\sqrt{20}} \leq \mu \leq 14,95 + 1,555 \frac{0,1}{\sqrt{20}}$$

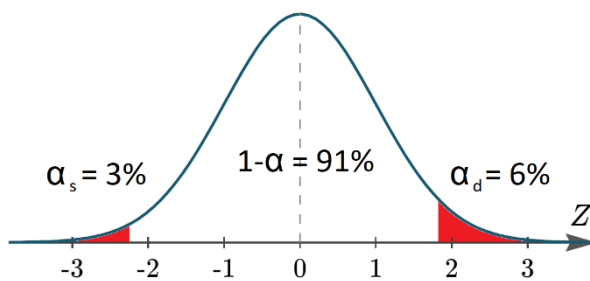


Fig. 4.5. Exemplu de calcul cu distribuția normală și RBA

Risc unilateral stânga, RUS

Se cunosc: n , \bar{x} , σ^2 , α .

Se determină: μ .

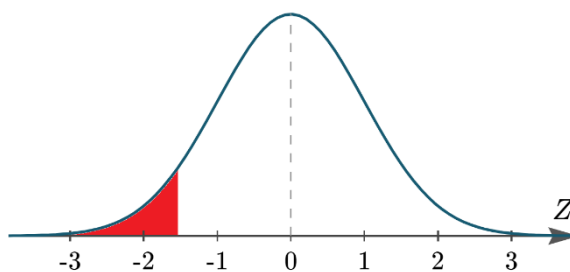


Fig. 4.6. Distribuția normală cu Risc Unilateral Stânga

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-z_\alpha \leq z) = 1 - \alpha$$

$$P = P\left(-z_\alpha \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

Din tabelul repartiției normale standard se alege valorile z_α . Riscul este doar în partea stângă (Figura 4.6). Din inegalitate se estimează intervalul în care se află media lotului:

$$\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} < \mu$$

Exemplu de calcul:

Se cunosc:

- $n = 25$;
- $\bar{x} = 12,05$;
- $\sigma^2 = 0,01$
- $\alpha = 4\%$; (RUS)

Se determină: $\mu = ?$

Pași de lucru:

$$P = P(z \geq -z_{\alpha}) = 96\%$$

$$P = P\left(\frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \geq -z_{\alpha}\right) = 96\%$$

Din tabelul repartiției normale $N(0,1) \Rightarrow z_{\alpha} = -1.75$

$$\mu \geq 12,05 - 1,75 \frac{0,1}{\sqrt{25}}$$

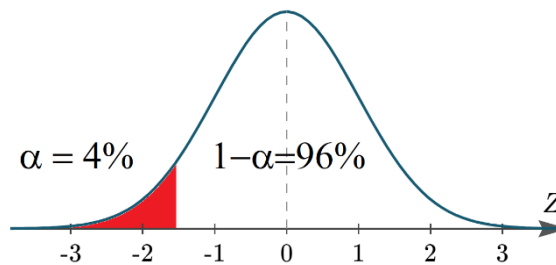


Fig. 4.7. Exemplu de calcul cu distribuția normală și RUS

Risc unilateral dreapta, RUD

Se cunosc: $n, \bar{x}, \sigma^2, \alpha$.

Se determină: μ .

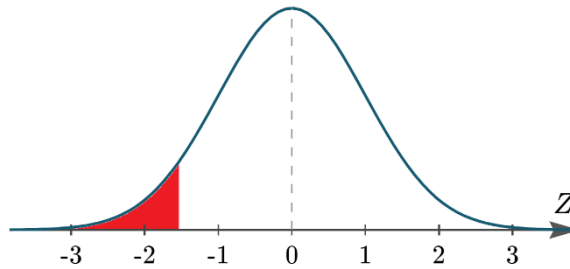


Fig. 4.8. Distribuția normală cu Risc Unilateral Dreapta

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(z \leq z_\alpha) = 1 - \alpha$$

$$P = P\left(\frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_\alpha\right) = 1 - \alpha$$

Din tabelul repartiției normale standard se alege valorile z_α ; . Riscul este doar în partea dreaptă (Figura 4.8). Din inegalitate se estimează intervalul în care se află media lotului:

$$\mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 12$;
- $\bar{x} = 8,5$;
- $\sigma^2 = 0,01$
- $\alpha = 8\%$ (RUD)

Se determină: $\mu = ?$

Pași de lucru:

$$P = P(z \leq z_\alpha) = 92\%$$

$$P = P\left(\frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_{\alpha}\right) = 92\%$$

Din tabelul repartiției normale $N(0,1) \Rightarrow z_{\alpha} = 1.405$

$$\mu \leq 8,5 + 1,405 \frac{0,1}{\sqrt{12}}$$

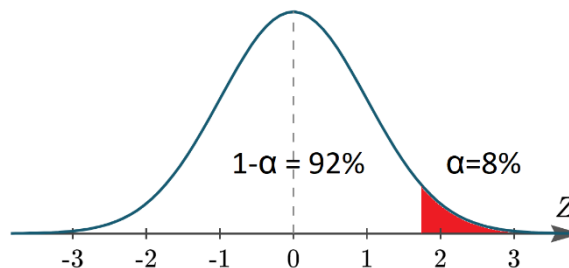


Fig. 4.9. Exemplu de calcul cu distribuția normală și RUD

4.2.1.2. Estimarea mediei când dispersia este necunoscută

Dintr-un lot a cărui dispersie σ^2 este necunoscută, (cazul beneficiarului), se extrage un eșantion de volum n pentru care se calculează media \bar{x} și dispersia eșantionului s^2 . Cu riscul α se estimează media lotului μ .

Statistica utilizată este:

$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}}$$

care conduce la o repartiție Student cu $v=n-1$ grade de libertate.

Valorile t se citesc din tabelul distribuției Student (Anexa 3)

Risc bilateral simetric, RBS

Se cunosc: n, \bar{x}, s^2, α .

Se determină: μ .

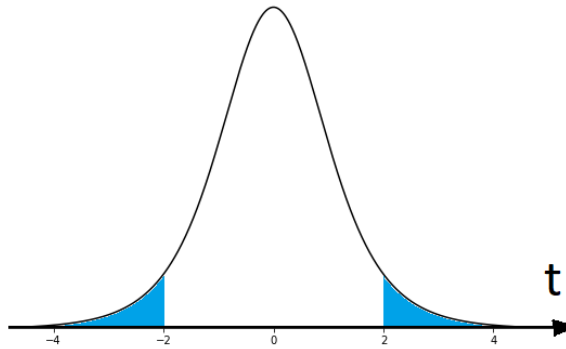


Fig. 4.10. Distribuția Student cu Risc Bilateral Simetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha$$

$$P = P\left(-t_{\alpha/2} \leq \frac{\mu - \bar{x}}{s/\sqrt{n}} \leq t_{\alpha/2}\right) = 1 - \alpha$$

Din tabelul repartiției Student se aleg valorile $\pm t_{\alpha/2}$

Din inegalitate se estimează intervalul în care se află media lotului:

$$\mu = \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 16; v=15$
- $\bar{x} = 10,1;$
- $s^2 = 0,01;$
- $\alpha = 10\%$ (RBS).

Se determină : $\mu = ?$

$$P = P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 90\%$$

$$P = P(-t_{\alpha/2} \leq \frac{\mu - \bar{x}}{s/\sqrt{n}} \leq t_{\alpha/2}) = 90\%$$

Din tabelul repartiției Student $\Rightarrow \pm t_{\alpha/2} = 1,753$

$$\mu = 10,1 \pm 1,753 \frac{0,1}{\sqrt{16}}$$

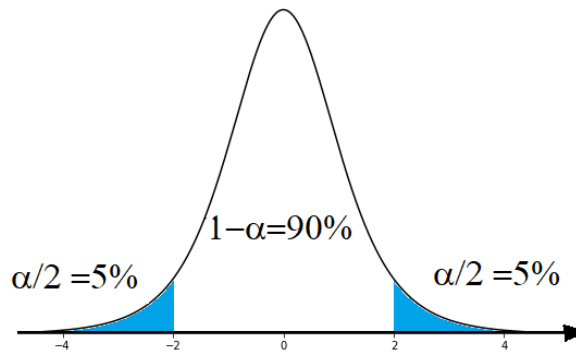


Fig. 4.11. Exemplu de calcul cu distribuția Student și RBS

Risc bilateral asimetric, RBA

Se cunosc: n, \bar{x}, s^2, α .

Se determină: μ .

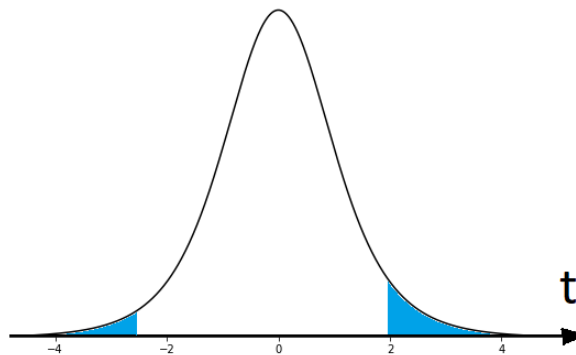


Fig. 4.12. Distribuția Student cu Risc Bilateral Asimetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-t_s \leq t \leq t_d) = 1 - \alpha$$

$$P = P\left(-t_s \leq \frac{\mu - \bar{x}}{s/\sqrt{n}} \leq t_d\right) = 1 - \alpha$$

Din tabelul repartiției Student se aleg valorile $-t_s$; t_d

Din inegalitate se estimează intervalul în care se află media lotului:

$$\bar{x} - t_s \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_d \frac{s}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 20$; $v=19$;
- $\bar{x} = 14,95$;
- $s^2 = 0,01$
- $\alpha_s = 5\%$; $\alpha_d = 10\%$ (RBA)

Se determină : $\mu = ?$

$$P = P(-t_s \leq t \leq t_d) = 85\%$$

$$P = P\left(-t_s \leq \frac{\mu - \bar{x}}{S/\sqrt{n}} \leq t_d\right) = 85\%$$

Din tabelul repartiției Student $\Rightarrow -t_s = -1,729$; $t_d = 1,328$

$$14,95 - 1,729 \frac{0,1}{\sqrt{20}} \leq \mu \leq 14,95 + 1,328 \frac{0,1}{\sqrt{20}}$$

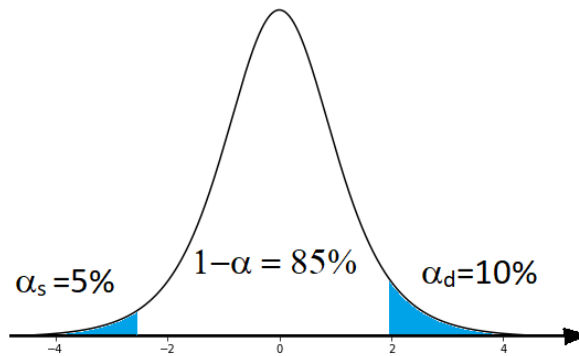


Fig. 4.13. Exemplu de calcul cu distribuția Student și RBA

Risc unilateral stânga, RUS

Se cunosc: n , \bar{x} , s^2 , α .

Se determină: μ .

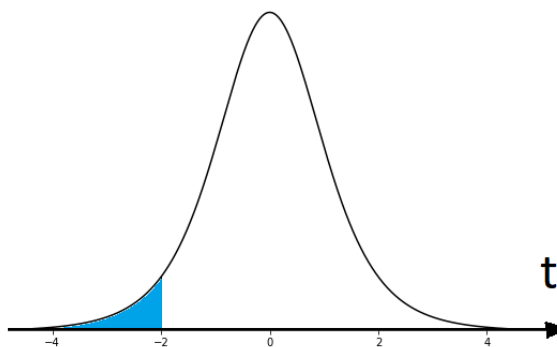


Fig. 4.14. Distribuția Student cu Risc Unilateral Stânga

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(-t_\alpha \leq t) = 1 - \alpha$$

$$P = P\left(-t_\alpha \leq \frac{\mu - \bar{x}}{s/\sqrt{n}}\right) = 1 - \alpha$$

Din tabelul repartiției Student se alege valoarea $-t_\alpha$

Din inegalitate se estimează intervalul în care se află media lotului:

$$\bar{x} - t_{\alpha} \frac{s}{\sqrt{n}} \leq \mu$$

Exemplu de calcul:

Se cunosc:

- $n = 25$; $v=24$;
- $\bar{x} = 12,05$;
- $s^2 = 0,01$
- $\alpha = 10\%$; (RUS))

Se determină : $\mu = ?$

$$P = P(t \geq -t_{\alpha}) = 90\%$$

$$P = P\left(\frac{\mu - \bar{x}}{s/\sqrt{n}} \geq -t_{\alpha}\right) = 90\%$$

Din tabelul repartiției Student $\Rightarrow -t_{\alpha} = -1,318$

$$\mu \geq 12,05 - 1,318 \frac{0,1}{\sqrt{25}}$$

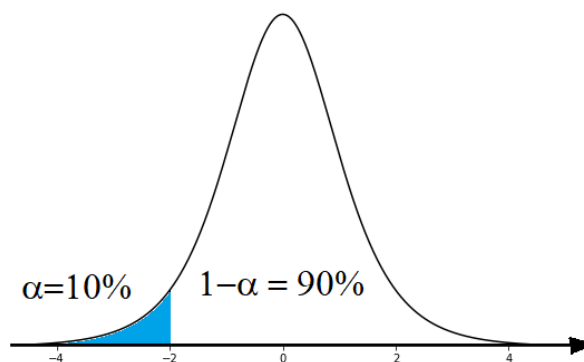


Fig. 4.15. Exemplu de calcul cu distribuția Student și RUS

Risc unilateral dreapta, RUD

Se cunosc: n , \bar{x} , s^2 , α .

Se determină: μ .

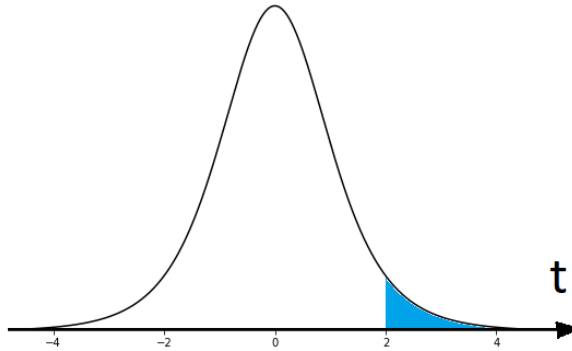


Fig. 4.16. Distribuția Student cu Risc Unilateral Dreapta

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(t \leq t_\alpha) = 1 - \alpha$$

$$P = P\left(\frac{\mu - \bar{x}}{s/\sqrt{n}} \leq t_\alpha\right) = 1 - \alpha$$

Din tabelul repartiției Student se alege valoarea t_α

Din inegalitate se estimează intervalul în care se află media lotului:

$$\mu \leq \bar{x} + t_\alpha \frac{s}{\sqrt{n}}$$

Exemplu de calcul:

Se cunosc:

- $n = 12$;
- $\bar{x} = 8,5$;
- $s^2 = 0,01$
- $\alpha = 5\%$ (RUD)

Se determină : $\mu = ?$

$$P = P(\leq t \leq t_\alpha) = 95\%$$

$$P = P\left(\frac{\mu - \bar{x}}{S/\sqrt{n}} \leq t_\alpha\right) = 95\%$$

Din tabelul repartiției Student $\Rightarrow -t_\alpha = 1,796$

$$\mu \leq 8,5 + 1,796 \frac{0,1}{\sqrt{12}}$$

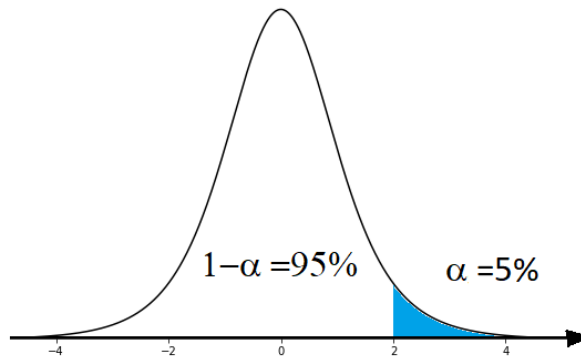


Fig. 4.17. Exemplu de calcul cu distribuția Student și RUD

4.2.2. Estimarea dispersiei

Dintr-un lot a cărui dispersie σ^2 este necunoscută, se extrage un eșantion de volum n pentru care se calculează dispersia eșantionului s^2 . Cu riscul α se estimează dispersia lotului σ^2 .

Statistica utilizată este:

$$\chi^2 = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = \frac{n-1}{\sigma^2} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} = (n-1) \frac{S^2}{\sigma^2}$$

care conduce la o repartiție χ^2 cu $v=n-1$ grade de libertate.

Valorile χ^2 se citesc din tabelul distribuției Chi-pătrat (Anexa 4)

Risc bilateral simetric, RBS

Se cunosc: n, \bar{x}, s^2, α .

Se determină: σ^2 .

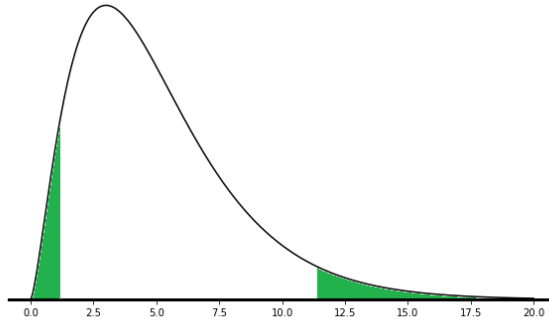


Fig. 4.18. Distribuția Chi-pătrat cu Risc Bilateral Simetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(\chi_{\alpha/2}^2 \leq \chi^2 \leq \chi_{1-\alpha/2}^2) = 1 - \alpha$$

$$P = P(\chi_{\alpha/2}^2 \leq (n-1) \frac{s^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2) = 1 - \alpha$$

Din tabelul repartiției Student se aleg valorile $\chi_{\alpha/2}^2$ și $\chi_{1-\alpha/2}^2$

Din inegalitate se estimează intervalul în care se află media lotului:

$$(n-1) \frac{s^2}{\chi_{1-\alpha/2}^2} \leq \sigma^2 \leq (n-1) \frac{s^2}{\chi_{\alpha/2}^2}$$

Exemplu de calcul:

Se cunosc:

- $n = 16; v=15$
- $S^2 = 0,01;$
- $\alpha = 10\%$ (RBS).

Se determină : $\sigma^2 = ?$

$$P = P(\chi_{\alpha/2}^2 \leq \chi^2 \leq \chi_{1-\alpha/2}^2) = 90\%$$

$$P = P(\chi_{\alpha/2}^2 \leq (n-1) \frac{s^2}{\sigma^2} \leq \chi_{1-\alpha/2}^2) = 90\%$$

Din tabelul repartiției $\chi^2 \Rightarrow \chi^2_{\alpha/2} = 24,996; \chi^2_{1-\alpha/2} = 7,261$.

$$15 \frac{0,01}{24,996} \leq \sigma^2 \leq 15 \frac{0,01}{7,261}$$

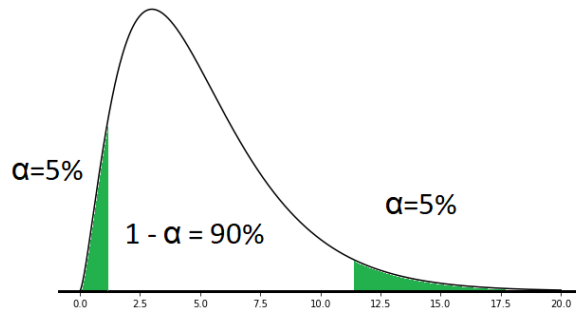


Fig. 4.19. Exemplu de calcul cu distribuția Chi-pătrat și RBS

Risc bilateral asimetric, RBA

Se cunosc: n, \bar{x}, s^2, α .

Se determină: σ^2 .

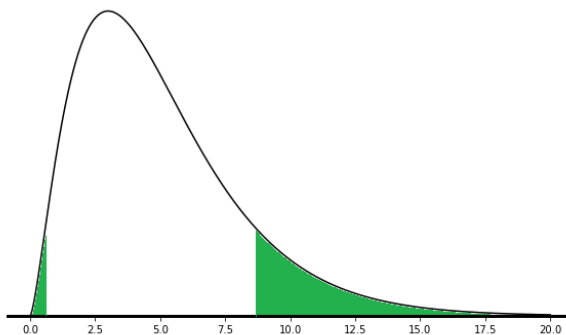


Fig. 4.20. Distribuția Chi-pătrat cu Risc Bilateral Asimetric

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(\chi_{\alpha_s}^2 \leq \chi^2 \leq \chi_{1-\alpha_d}^2) = 1 - \alpha$$

$$P = P(\chi_{\alpha_s}^2 \leq (n-1) \frac{s^2}{\sigma^2} \leq \chi_{1-\alpha_d}^2) = 1 - \alpha$$

Din tabelul repartiției Student se aleg valorile $\chi_{\alpha_s}^2$ și $\chi_{1-\alpha_d}^2$

Din inegalitate se estimează intervalul în care se află media lotului:

$$(n-1) \frac{s^2}{\chi_{1-\alpha_d}^2} \leq \sigma^2 \leq (n-1) \frac{s^2}{\chi_{\alpha_s}^2}$$

Exemplu de calcul:

Se cunosc:

- $n = 20$; $v=19$;
- $\sigma^2 = 0,01$
- $\alpha_s = 5\%$; $\alpha_d = 10\%$ (RBA).

Se determină : $\sigma^2 = ?$

$$P = P(\chi_{\alpha_s}^2 \leq \chi^2 \leq \chi_{\alpha_d}^2) = 85\%$$

$$P = P(\chi_{\alpha_s}^2 \leq n-1 \frac{s^2}{\sigma^2} \leq \chi_{\alpha_d}^2) = 85\%$$

Din tabelul repartiției $\chi^2 \Rightarrow \chi_{\alpha_s}^2 = 10,117$ $\chi_{1-\alpha_d}^2 = 27,204$

$$19 \frac{0,01}{27,204} \leq \sigma^2 \leq 19 \frac{0,01}{10,117}$$

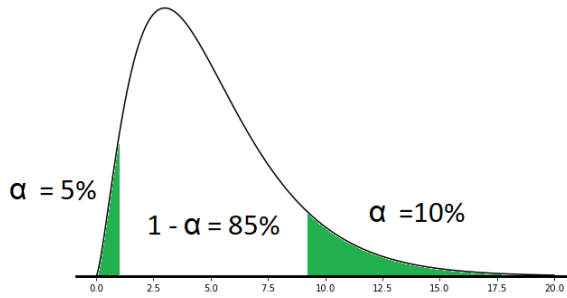


Fig. 4.21. Exemplu de calcul cu distribuția Chi-pătrat și RBA

Risc unilateral stânga, RUS

Se cunosc: n, \bar{x}, s^2, α .

Se determină: σ^2 .

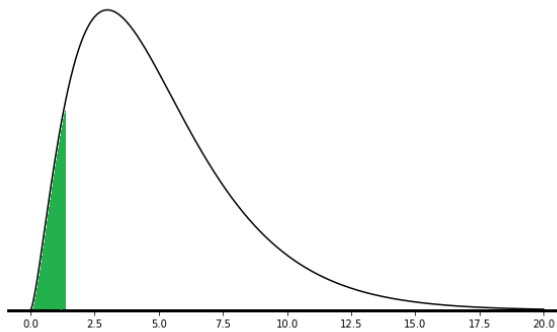


Fig. 4.22. Distribuția Chi-pătrat cu Risc Unilateral Stânga

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(\chi_{1-\alpha}^2 \geq \chi^2) = 1 - \alpha$$

$$P = P(\chi_{1-\alpha}^2 \geq (n - 1) \frac{s^2}{\sigma^2}) = 1 - \alpha$$

Din tabelul repartiției Student se alege valoarea χ_{α}^2

Din inegalitate se estimează intervalul în care se află media lotului:

$$(n - 1) \frac{s^2}{\chi_{1-\alpha}^2} \geq \sigma^2$$

Exemplu de calcul:

Se cunosc:

- $n = 25$; $v=24$;
- $\bar{x} = 12,05$;
- $\sigma^2 = 0,01$;
- $\alpha = 10\%$; (RUS)

Se determină : $\sigma^2 = ?$

$$P = P(\chi_{1-\alpha}^2 \geq \chi^2) = 90\%$$

$$P = P(\chi_{1-\alpha}^2 \geq (n - 1) \frac{s^2}{\sigma^2}) = 90\%$$

Din tabelul repartiției $\chi^2 \Rightarrow \chi^2_{1-\alpha} = 15,659$;

$$\sigma^2 \leq 24 \frac{0,01}{15,659}$$

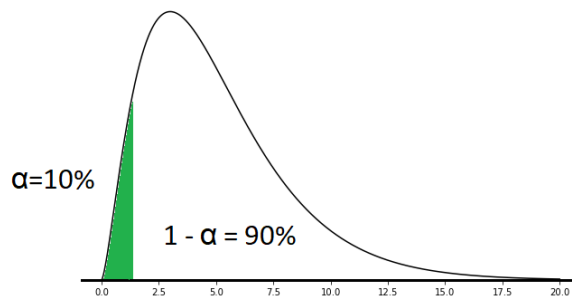


Fig. 4.23. Exemplu de calcul cu distribuția Chi-pătrat și RUS

Risc unilateral dreapta, RUD

Se cunosc: n, \bar{x}, s^2, α .

Se determină: σ^2 .

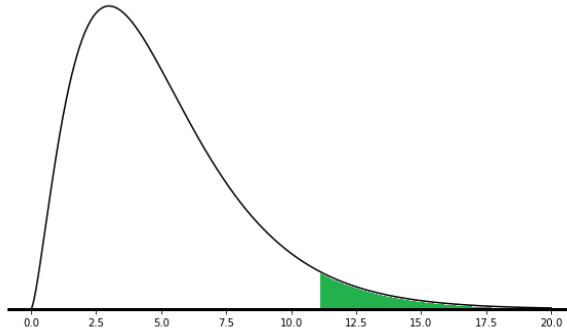


Fig. 4.24. Distribuția Chi-pătrat cu Risc Unilateral Dreapta

Relația care leagă intervalul de încredere cu nivelul de încredere:

$$P = P(\chi^2 \leq \chi_\alpha^2) = 1 - \alpha$$

$$P = P\left((n - 1) \frac{s^2}{\sigma^2} \leq \chi_\alpha^2\right) = 1 - \alpha$$

Din tabelul repartiției Student se alege valoarea χ_α^2

Din inegalitate se estimează intervalul în care se află media lotului:

$$(n - 1) \frac{s^2}{\chi_\alpha^2} \leq \sigma^2$$

Exemplu de calcul:

Se cunosc:

- $n = 16; v=15$
- $S^2 = 0,01;$
- $\alpha = 10\%$ (RBS).

Se determină : $\sigma^2 = ?$

$$P = P(\chi^2 \leq \chi^2_\alpha) = 1 - \alpha$$

$$P = P\left((n - 1) \frac{s^2}{\sigma^2} \leq \chi^2_\alpha\right) = 1 - \alpha$$

Din tabelul repartiției $\chi^2 \Rightarrow \chi^2_\alpha = 19,675$

$$11 \frac{0,01}{19,675} \leq \sigma^2$$

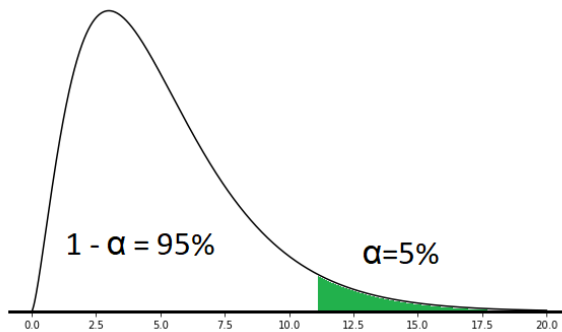


Fig. 4.25. Exemplu de calcul cu distribuția Chi-pătrat și RUD

4.3. Descrierea temei

Lucrați într-o firmă care face piese prin prelucrări mecanice. Trebuie să estimați parametrii unor loturi de piese. Pentru primele 4 loturi (Lot 1 până la Lot 4) cunoașteți abaterea standard a populației $\sigma=0.16$ iar pentru restul abaterea standard a eșantionului este $s=0.XX$ unde XX este ziua Dvs. de naștere. Pentru celelalte loturi (Lot 5 până la Lot 12) parametrii populației sunt necunoscuți. Din fiecare lot eșantionați câte 30 de valori.

Trebuie să generați un raport cu următoarele informații:

- Pentru primele patru loturi (Lot 1 până la Lot 4) trebuie să estimați media cu diferitele nivele de risc din Tab 4.1.: Risc Unilateral Stânga (RUS), Risc Unilateral Dreapta (RUD), Risc Bilateral Simetric (RBS) și Risc Bilateral Asimetric (RBA) ;

- Pentru următoarele patru loturi (Lot 5 până la Lot 8) estimați media știind că dispersia lotului este necunoscută pentru nivelele de risc (RUS, RUD, RBS, RBA) din Tab. 4.1. Folosiți dispersia eșantionului pentru estimare;
- Pentru ultimele patru loturi (Lot 9 până la Lot 12) estimați dispersia folosind dispersia eșantionului pentru cele patru nivele de risc (RUS, RUD, RBS, RBA) precizate în Tab. 4.1.

Problema poate fi prezentată sintetic sub forma tabelului Tab. 4.1.

Tipul problemei:	Estimarea Mediei										Estimarea dispersiei				
	Se cunoaște dispersia					Nu se cunoaște dispersia					Nu se cunoaște dispersia				
Tipul riscului:	RUS	RUD	RBS	RBA		RUS	RUD	RBS	RBA		RUS	RUD	RBS	RBA	
Riscul	3%	6%	8%	2%	3%	5%	1%	10%	5%	2%	2%	10%	10%	5%	2.5%
Eșantioane din:	Lot 1	Lot 2	Lot 3	Lot 4		Lot 5	Lot 6	Lot 7	Lot 8		Lot 9	Lot 10	Lot 11	Lot 12	

Pentru generarea raportului puteți urma pașii:

- **Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din primele patru loturi (Lot 1 până la Lot 4). Șirul va avea media 20 și abaterea standard 0.16;**

Generarea numerelor aleatorii a fost prezentată în secțiunea [1.2.1. Generarea unui șir de numere aleatorii.](#)

- **Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din loturile de la Lot 5 la Lot 8. Șirul va avea media 12 și abaterea standard 0.XX unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15);**

- **Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din ultimele patru loturi. Șirul va avea media 35 și abaterea standard 0.XX unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15);**

- **Calculați media și abaterea standard a fiecărui eșantion a fiecărui lot.**

Modul de calcul al indicatorilor este prezentat în secțiunile [2.2.1. Indicatori de localizare](#) și [2.2.2. Indicatori de împrăștiere](#).

- **Determinați, pentru fiecare nivel de risc, statistica corespunzătoare (z, t sau χ^2)**

Modul de determinare a fiecărei statistici este prezentat în secțiunile [4.2.1. Estimarea mediei](#) și [4.2.2. Estimarea dispersiei](#).

Tabelele repartițiilor se găsesc în *Anexa 1*.

- **Calculați capătul sau capetele intervalelor de încredere pentru fiecare lot în parte.**

Modul de calcul este prezentat în secțiunile [4.2.1. Estimarea mediei](#) și [4.2.2. Estimarea dispersiei](#).

- **Generați un raport cu valorile obținute după model.**
- **BONUS: Estimați media și dispersia pentru un eșantion de 20 valori normal distribuite cu media de 30 și abaterea standard de 0.5 astfel:**
 - RBS, RBA pentru estimarea mediei când $\sigma = 0.3$
 - RUS, RBS pentru estimarea mediei când σ este necunoscută
 - RBA, RUD pentru estimarea dispersiei.

Sarcină	Unde trebuie să apară
a. Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din primele patru loturi (Lot 1 până la Lot 4). Șirul va avea media 20 și abaterea standard 0.16;	Foaie de lucru
b. Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din loturile de la Lot 5 la Lot 8. Șirul va avea media 12 și abaterea standard 0.XX unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15);	Foaie de lucru
c. Generați un șir pe 4 coloane cu câte 30 de valori normal distribuite fiecare reprezentând eșantioanele din ultimele patru loturi. Șirul va avea media 35 și abaterea standard 0.XX unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15);	Foaie de lucru
d. Calculați media și abaterea standard a fiecărui eșantion a fiecărui lot.	Raport
e. Determinați, pentru fiecare nivel de risc, statistica corespunzătoare (z, t sau χ^2)	Raport
f. Calculați capătul sau capetele intervalelor de încredere pentru fiecare lot în parte.	Raport
g. Generați un raport cu valorile obținute după model.	

h. BONUS: Estimați diferența între media Lotului 1 și 2 și apoi între 3 și 4	Raport
--	--------

4.4. Resurse

Descriere resursa	Limba	Format	Link
Estimarea mediei	EN	video	https://www.youtube.com/watch?v=x6OsGXwi1hU
Estimarea mediei (Prezentare PPT)	RO	text	http://www.mpopa.ro/statistica-master/01_inferenta_estimarea.pdf
Estimarea parametrilor (pag 36)	RO	text	http://web-ng.info.uvt.ro/~balint/files/statistica.pdf

5. Determinarea capacității și crearea unei cartele de control

5.1. [Obiective de învățare](#)

5.2. [Concepte teoretice](#)

5.3. [Descrierea temei](#)

5.4. [Resurse](#)

5.1. Obiective de învățare

În această lucrare veți învăța despre controlul statistic al proceselor, structura unei cartele de control și modul în care se calculează capacitatea unui proces.

De asemenea veți învăța cum să determinați coeficientul de corelație Pearson și să trasați o linie de regresie.

5.2. Concepte teoretice

În epoca modernă a ingineriei și a producției, capacitatea de a menține o calitate constantă și de a minimiza variabilitatea proceselor nu este doar ideală, este esențială. Aici se folosește controlul statistic al proceselor (SPC) ca o metodologie foarte importantă. SPC este o metodologie științifică, bazată pe date, pentru monitorizarea, controlul și îmbunătățirea procedurilor și a rezultatelor în procesele de fabricație și de afaceri. Este un segment al domeniului mai larg al managementului calității care utilizează în mod specific metodele statistice pentru a-și atinge obiectivele.

SPC se bazează pe premisa că fiecare proces are variabilitate. Această variabilitate se poate datora unor cauze comune care sunt inerente procesului sau unor cauze speciale care nu fac parte din arhitectura procesului.

Înțelegerea și controlul acestei variabilități sunt caracteristici importante în obținerea unui produs sau serviciu de calitate superioară.

Obiectivele primare ale SPC sunt:

1. Să monitorizeze comportamentul procesului: Prin SPC, veți învăța să creați cartele de control care oferă o reprezentare vizuală a unui proces în timp. Aceste diagrame ajută la identificarea tendințelor, modelelor și variațiilor.
2. Controlul variabilității procesului: Făcând distincția între variabilitatea de cauză comună și variabilitatea de cauză specială, puteți controla și reduce în mod eficient variabilitatea globală a procesului.
3. Îmbunătățirea capacităților procesului: SPC nu se referă doar la menținerea status quo-ului, ci și la îmbunătățirea continuă. Prin intermediul diferitelor instrumente SPC, puteți identifica oportunități de îmbunătățire a capacităților și performanțelor procesului.

Instrumente și tehnici SPC

De-a lungul acestui capitol, vom aborda o serie de instrumente și tehnici asociate cu SPC, inclusiv:

- Cartele de control: aceste diagrame sunt utilizate pentru monitorizarea stabilității procesului și identificarea variațiilor care necesită acțiune.
- Analiza capabilității procesului: Aceasta implică calcularea indicilor de capacitate, cum ar fi C_p sau C_{pk} pentru a evalua cât de bine un proces îndeplinește specificațiile.
- Diagrame de cauză și efect: Cunoscute și sub numele de diagrame pește sau diagrame Ishikawa, acestea ajută la identificarea cauzelor potențiale ale variabilității procesului.
- Analiza Pareto: O tehnică de prioritizare a cauzelor potențiale ale problemelor pentru a le aborda pe cele mai semnificative.

- Histograme: Acestea oferă un rezumat vizual al variației datelor și pot indica dacă ieșirile procesului sunt distribuite într-un mod care respectă cerințele de calitate.

Controlul statistic se utilizează în domeniul producției de serie mijlocie până la mare, pentru:

- fabricarea produselor materiale cum ar fi:
 - piese pentru mașini;
 - produse chimice;
 - produse de turnătorie;
 - piese din materiale plastice;
 - subansamble electrice;
 - medicamente;
 - ambalaje;
- supravegherea caracteristicilor procesului sau produsului:
 - dimensiunile particulelor;
 - valoarea pH;
 - concentrația;
 - gradul de puritate;
- supravegherea produselor și serviciilor:
 - software;
 - servicii bancare;
 - servicii telefonice;
 - servicii de livrare.

Avantajele care pot fi obținute prin utilizarea metodei SPC sunt:

- evitarea erorilor în producție;
- reducerea măsurilor de verificare în controlul final;
- posibilitatea de detectare a erorilor care nu mai pot fi evidențiate pe standul de verificare finală;

- supravegherea procesului de fabricație și garantarea că etapele următoare ale procesului nu vor conține nici o piesă defectă din punct de vedere al dimensiunilor inițiale de intrare (procesul următor este considerat clientul beneficiar al procesului anterior);
- detectarea și eliminarea mărimilor perturbatoare ale unui proces atât în ceea ce privește amplitudinea acestora cât și în ceea ce privește optimizarea parametrilor care influențează procesul, cum sunt de exemplu materialul și toleranțele piesei, specificațiile referitoare la mașini, reglarea sculei sau specificațiile referitoare la mijloacele de verificare;
- identificarea timpurie a problemelor referitoare la calitate, procesele tehnologice;
- stabilitatea fabricației, adică menținerea sub control statistic a tuturor proceselor de fabricație
- reducerea costurilor, a procentului de rebuturi și a cheltuielilor de verificare atât în ceea ce privește numărul lor cât și restrângerea sferei de verificare
- creșterea nivelului calitativ general și ca urmare creșterea productivității prin utilizarea sistematică a analizelor și prin documentarea lor, prin evaluarea prognozelor pe termen lung și printr-un procedeu continuu de feedback aplicat datelor de măsurare.

Caracteristicile verificate se pot diviza în două categorii importante care la rândul lor pot fi divizate în:

1. Caracteristici cantitative, care pot fi măsurate:

1.1 Caracteristici cantitative continue (exprimate prin valori care aparțin mulțimii numerelor reale):

- temperatura băii de lipire, în [°K];
- masa de umplere a unui ambalaj, în [g];

- diametrul unei piese aflate în mișcare de rotație, în [mm];
- fiabilitatea unui bec cu incandescență, în [h];

1.2 Caracteristici cantitative discrete continue (exprimate prin valori care aparțin mulțimii numerelor întregi):

- numărul valorilor de măsurare aflate domeniul de toleranță;
- numărul pieselor din lot care prezintă defecte;
- numărul erorilor de tipar de pe fiecare pagină a unei cărți;

2. Caracteristici atributive (exprimate prin atribute de genul corect-defect):

2.1 Caracteristici atributive ordinale:

- Temperatura (Rece, Călduț, Cald, Fierbinte)
- Calificative școlare ("foarte bine", "bine", "satisfăcător")

2.2 Caracteristici atributive nominale:

- Culoarea figurinelor de jucărie (albastru, galben, roșu)
- Fixarea prețului de cost (646, 689, 693)

Există două tipuri diferite de factori de influențare care acționează asupra unui proces:

- factorii de influențare accidentală (aleatorii). Cauzele apariției acestora pot fi minore și nu pot fi prevăzute. Apariția lor se face neregulat fiind oricând posibilă apariția lor dacă nu se iau măsuri de înlăturare.
- factorii de influențare sistematică. Apariția lor se datorează unor cauze importante previzibile prin aplicarea teoriei probabilităților.

Pentru supravegherea mărimilor cantitative ale unui proces se utilizează diverse funcții de distribuție. În cazul caracteristicilor atributive (bun/rău) se va utiliza adesea distribuția binomială, respectiv distribuția Poisson. Caracteristicile variabile (măsurabile) au o distribuție normală în majoritatea cazurilor. La realizarea unei cuprinzătoare examinări preliminare înainte de aplicarea SPC, prin procedeele statistice de testare trebuie să se verifice dacă

distribuția reală a valorilor caracteristice poate fi apreciată ca fiind asemănătoare cu una dintre distribuțiile teoretice menționate mai sus.

Pentru aplicarea controlului statistic trebuie verificat dacă procesele sunt controlabile și performante. Un proces este considerat controlabil atunci când distribuția caracteristicilor procesului se menține practic nemodificată, respectiv se modifică numai în limite cunoscute. Un proces este considerat performant dacă este capabil să furnizeze unități care îndeplinesc cerințele de calitate, mai precis, dacă numărul rebuturilor rezultate din proces este practic aproape nul.

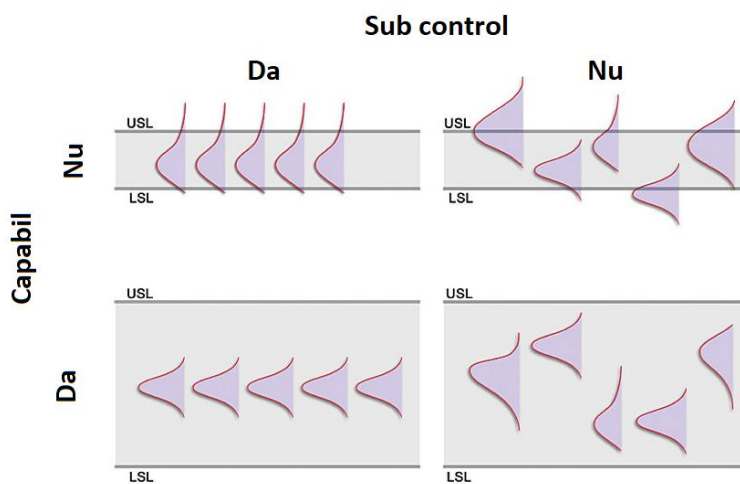


Fig. 5.1 Capacitatea proceselor de a fi controlabile și performante

Sursa: "10 Keys for Maximizing the Benefits of your SPC Program | 2016-03-01 | Quality Magazine." Accesat: Nov. 08, 2023. [Online]. Disponibil: <https://www.qualitymag.com/articles/93141-keys-for-maximizing-the-benefits-of-your-spc-program>

Exemplul prezentate în Fig.5.1 (jos stânga) este a unui proces controlabil, deoarece abaterile mediile aritmetice sunt minime și performant deoarece curbele de repartiție sunt toate ascuțite cea ce denotă o dispersie mică. În partea sus stânga este a unui proces controlabil dar neperformant (curbele de repartiție sunt aplatizate, fapt datorat unor erori sistematice). Prin înlăturarea acestora procesul poate fi controlabil statistic. În partea dreapta jos se prezintă cazul unui proces performant dar necontrolabil fapt datorat variației mari a mediilor eșantioanelor. Cauza este apariția unor abateri aleatorii care

trebuie identificate și îndepărtate. Ultimul caz prezentat în partea dreapta sus este al unui proces necontrolabil și neperformant, caz în care nu se recomandă aplicarea controlului statistic.

5.2.1. Capabilitate

Pentru obținerea unei aprecieri cantitativă a capacității performante a unui proces tehnologic, se calculează indici de performanță dintre care cei mai importanți sunt capabilitatea procesului și capabilitatea mașini, Calculul capabilității se efectuează în condiții reale de producție. Cunoașterea capacităților performante a unui proces tehnologic ne ajută să apreciem dacă procesul este controlabil pe durată îndelungată și dacă asigură calitatea necesară. Capabilitatea unui proces, C_p , C_{pk} și capabilitatea unei mașini, C_m , C_{mk} depind de poziția și lățimea repartiției față de limitele de toleranță (LST - limita superioară de toleranță; LIT - limita inferioară de toleranță) care pot fi simetrice sau asimetrice față de valoarea nominală. Calcularea celor doi indici rezultă din tabelul 5.1.

Variația capabilității se datorează deplasării mediei sau aplatizării curbelor de repartiție. Este de remarcă că industria auto din Germania impune capabilității de minim 2,5 în timp ce industria auto japoneză impune valori de 3,00.

Tabel 5.1. Calculul capabilității proceselor și mașinilor

C_m, C_{mk}	Capacitățile performante ale mașinii		
C_p, C_{pk}	Capacitățile performante ale procesului		
Capacitățile performante ale mașinii		Capacitățile performante ale procesului	
$C_m = \frac{LST - LIT}{6s}$		$C_p = \frac{LST - LIT}{6s}$	
$C_{ms} = \frac{LST - \bar{x}}{3s}$	$C_{mi} = \frac{\bar{x} - LIT}{3s}$	$C_{ps} = \frac{LST - \bar{x}}{3s}$	$C_{pi} = \frac{\bar{x} - LIT}{3s}$
$C_{pk} = \text{Min}(C_{mi}, C_{ms})$		$C_{pk} = \text{Min}(C_{pi}, C_{ps})$	

$C_m, C_p > 1,33$	Procesul este performant din punct de vedere calitativ
$1,00 \leq C_m, C_p \leq 1,33$	Procesul poate deveni performant în anumite condiții, Supraveghere atentă
$C_m, C_p < 1,00$	Procesul nu este performant

La examinarea capacităților performante ale unui proces se ține cont de diferiții factori care influențează procesul. Cunoașterea capacității performante a mașinii (și a actualei capacități performante a procesului) ne ajută să apreciem dacă procesul poate să îndeplinească cerințele de calitate impuse, ne ajută de asemenea în aprecierea unor procedee sau a unor mașini noi sau modificate și în colectarea datelor pentru întocmirea cartelelor de calitate.

Pentru introducerea cu succes a SPC trebuie să fie îndeplinite diferite premise:

- Procesul este controlabil;
- Procesul se află sub control statistic;
- Se cunosc caracteristicile produsului;
- Se cunoaște forma distribuției;
- Procesul este performant;
- Mijloacele de măsurare sunt performante;
- Angajații sunt instruiți și motivați;

5.2.2. Structura generală a cartelelor de control a calității

Structura de principiu a unei cartele este reprezentată în figura 5.2. Pe cartela de control pot fi indicate limitele de avertizare, limitele de intervenție și limitele de toleranță. Cartelele permit controlul continuu al procesului ceea ce este în beneficiul procesului. Este posibil astfel să se obțină o producție constantă și previzibilă atât în ceea ce privește costurile cât și în ceea ce

privește calitatea. Se constată o reducere a dispersiei produselor, reducerea costurilor și creșterea capacității efective. Cartelele de controlul a calității reprezintă o formă standardizată pentru schimbul de informații cu privire la descrierea calității procesului.

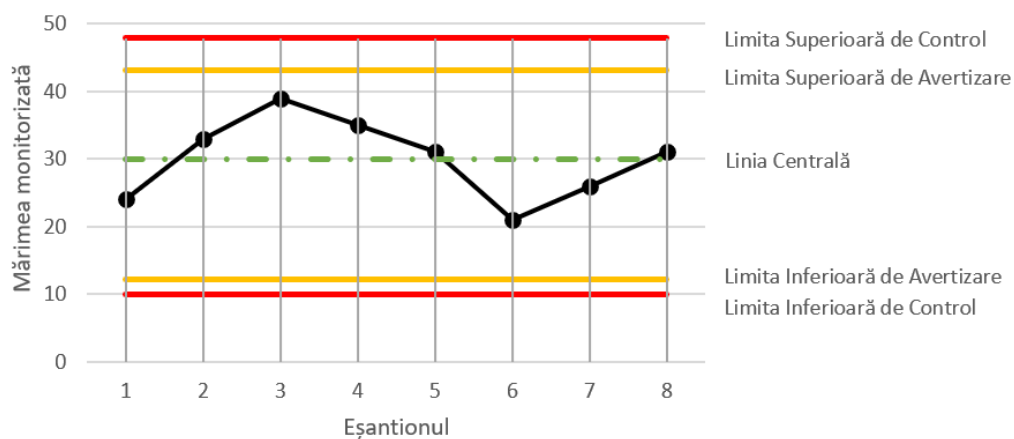


Fig. 5.2 Elementele componente ale cartelelor de control statistic

Pentru completarea cartelelor de controlul calității este necesar ca întreprinderile să definească principiile care trebuie respectate de către angajați, de exemplu:

- Înregistrările se vor efectua cu regularitate, conform planului de verificare;
- Se vor nota valorile inițiale, pentru a permite verificarea calculelor;
- Completarea se va efectua cu exactitate, pentru evitarea erorilor;
- Se vor uni punctele de pe cartela de control;
- Pentru a permite identificarea cauzelor erorilor, în fișa de însoțire a procesului trebuie înregistrate toate modificările și influențele asupra procesului.

La interpretarea unei cartele de control trebuie să urmărim o serie de reguli, prezentate în tabelul 5.2.

Tabel 5.2. Detecția problemelor într-un proces cu ajutorul cartelei de control

Regula	Descriere
1. Puncte înafara limitelor	Unul sau mai multe puncte sunt dincolo de limite
2. Test Zona A	2 din 3 puncte consecutive sunt în zona A sau mai departe
3. Test Zona B	4 din 5 puncte consecutive sunt în zona B sau mai departe
4. Test Zona C	7 sau mai multe puncte consecutive sunt de o singură parte a mediei (în Zona C sau mai departe)
5. Tendință	7 puncte consecutive au o tendință în sus sau în jos
6. Amestecare	8 puncte consecutive fără nici un punct în zona C
7. Stratificare	15 puncte consecutive în zona C
8. Supra-control	14 puncte consecutive alternând

Aceste reguli sunt ilustrate grafic în Figura 5.3.

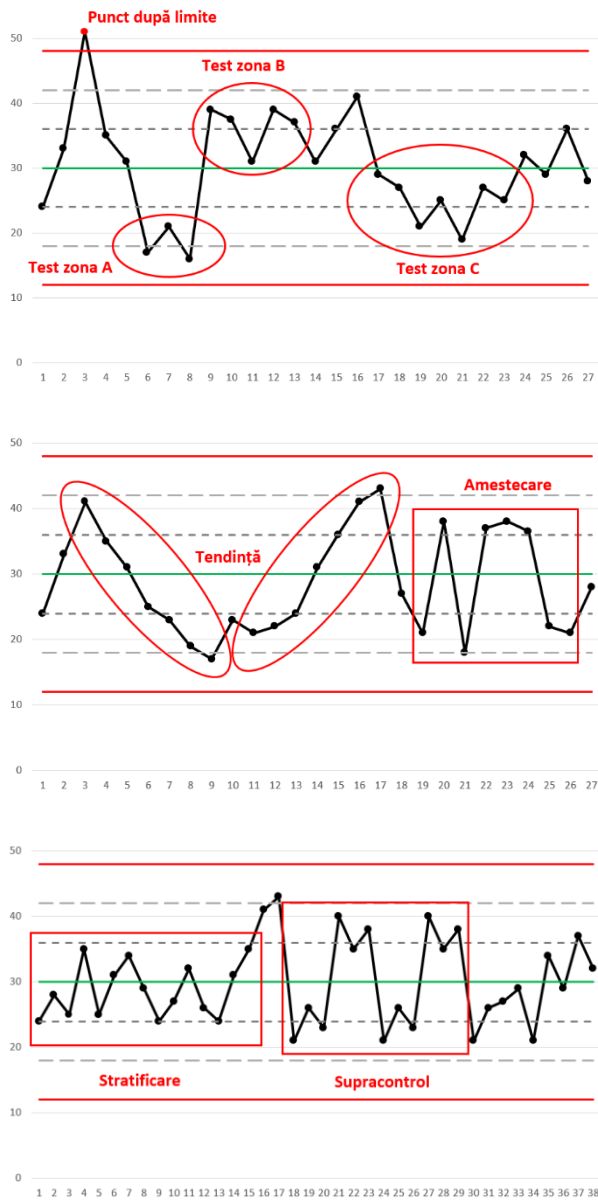


Fig. 5.3 Interpretarea cartelelor de control

5.2.3. Corelația

Corelația Pearson este utilizată pentru a determina puterea și direcția unei relații liniare între două variabile continue. Mai precis, testul generează un coeficient numit **coeficient de corelație Pearson**, notat ca r . Valoarea sa poate varia de la -1 pentru o relație liniară negativă perfectă la +1 pentru o relație

liniară pozitivă perfectă (Figura 5.4). O valoare 0 (zero) indică că nu există nicio relație între cele două variabile.

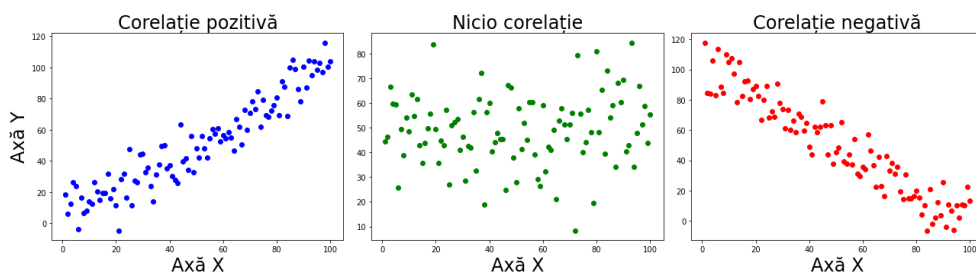


Fig. 5.4. Tipuri de corelații

Pentru calculul corect al coeficientului de corelație trebuie îndeplinite niște condiții:

- Nivelul de măsură trebuie să fie interval sau rație
- Variabilele trebuie să fie aproximativ normal distribuite
- Asocierea între cele două variabile trebui să fie liniară
- Datele ar trebui să nu conțină valori aberante

Relația de calcul a coeficientului de corelație este:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

unde:

x_i – valorile primei variabile

y_i – valorile celei de-a doua variabile

\bar{x} – media valorilor primei variabile

\bar{y} – media valorilor celei de-a doua variabile

5.2.4. Regresia liniară

Regresia liniară simplă evaluează relația liniară dintre două variabile continue pentru a prezice valoarea unei variabile dependente pe baza valorii unei variabile independente. Mai precis, ne ajută să:

- determinăm dacă relația liniară dintre două variabile este semnificativă statistic,
- determinăm cât din variația variabilei dependente este explicată de variabila independentă,
- înțelegem direcția și oricărei relații,
- prezicem valorile variabilelor dependente pe baza valorilor diferite ale variabilei independente.

Ideea de bază este de a obține o dreaptă care se potrivește cel mai bine datelor. Dreapta care se potrivește cel mai bine este cea pentru care eroarea totală de predicție este cât mai mică posibil. Eroarea este distanța dintre punct și linia de regresie.

Dreapta de regresie are ecuația:

$$Y = a * X + b$$

unde:

Y – variabila dependentă

X – variabila independentă

a, b – parametrii dreptei de regresie

Eroare de predicție se calculează făcând suma pătratelor diferențelor dintre valorile observate (din setul de date) și valorile prezise:

$$\Delta = \sum (y - \hat{y})^2$$

unde:

y – valorile observate

\hat{y} – valorile prezise

Reprezentarea grafică a unei drepte de regresie pentru două variabile X și Y este reprezentată în Figura 5.9. Fiecare observație este reprezentată printr-un punct care are valorile variabilelor X și respectiv Y. Spre exemplu, punctul marcat cu roșu, are valorile X=30 și Y=30. Dreapta care se potrivește cel mai bine pe aceste date are ecuația:

$$Y = 1.1269 * X - 6.3076$$

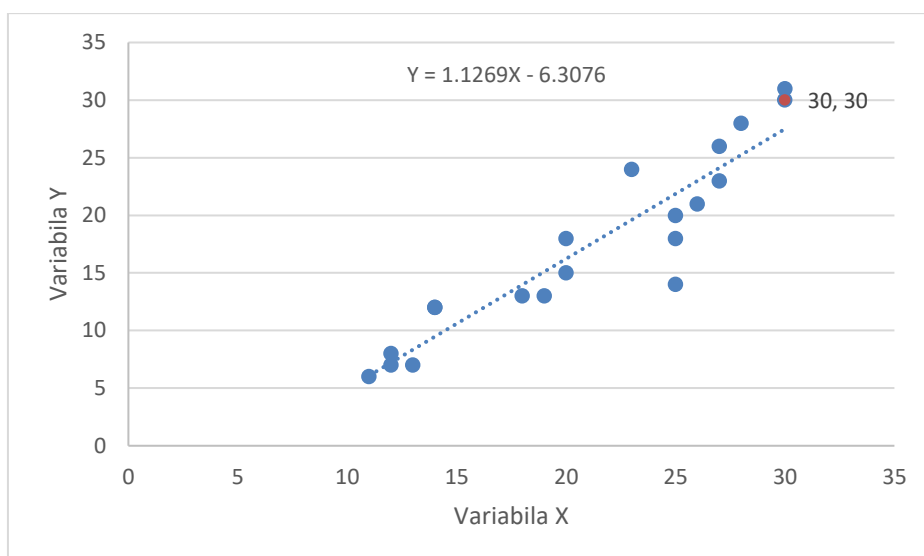


Fig.5.9 Dreapta de regresie pentru două variabile X și Y.

Modelul de regresie obținut poate fi evaluat cu ajutorul coeficientului de determinare (R^2). Acesta este o măsură statistică a cât de aproape sunt datele de linia de regresie găsită. R^2 reprezintă procentul variației variabilei dependente (de răspuns) explicat de modelul liniar:

$$R^2 = \frac{\text{Variație explicată}}{\text{Variație totală}}$$

Acest indicator ia valori între 0 și 100%, unde 0% indică faptul că modelul nu explică variabilitatea datelor de răspuns în jurul mediei sale iar 100% indică faptul că modelul explică toată variabilitatea datelor de răspuns în jurul valorii medii. În general, cu cât este mai mare R^2 , cu atât modelul se potrivește mai bine cu datele.

5.3. Descrierea temei

La firma în care lucrați s-a înființat o nouă linie de producție. Trebuie să determinați parametrii procesului, să creați o cartel de control și să calculați capabilitatea acestui proces. Eșantionați din producția de astăzi câte 5 piese în fiecare oră pentru 20 de ore. Creați cartela de control și tabelul frecvențelor și determinați parametrii de proces și capabilitatea. Trasați o histogramă și trageți concluzia.

Pentru a genera raportul puteți urma pașii:

- a. **Generați un șir de valori normal distribuite, pe 5 coloane a câte 20 de valori reprezentând cele 5 valori eșantionate în fiecare oră timp de 20 de ore. Media este 10 iar abaterea standard este egală cu $0.XX$ unde XX este ziua Dvs. de naștere (Exemplu: pentru cineva născut în data de 7 ian. abaterea standard va fi 0.07 iar pentru cineva născut în data de 15 mai abaterea standard va fi 0.15);**

Generarea unui șir de valori aleatorii este explicată în secțiunea [1.2.1. Generarea unui șir de numere aleatorii](#)

- b. **Calculați parametrii eșantionului: valorile minimă și maximă, amplitudinea, media aritmetică a mediilor eșantionului, dispersia și abaterea standard a eșantionului;**

Determinarea parametrilor statistici este explicată în secțiunea [2.2.1 Indicatori de localizare](#) și [2.2.2 Indicatori de împrăștiere](#) iar funcțiile folosite se găsesc în secțiunea [2.2.3. Funcții Excel pentru calculul indicatorilor statistici](#)

- c. **Împărțiți amplitudinea în 7 intervale egale și determinați frecvența absolută și relativă pentru fiecare din cele 7 intervale.**

Împărțirea pe intervale și calculul frecvențelor este detaliată în secțiunea [2.2.4. Noțiunea de frecvență și trasarea tabelului de frecvențe](#)

- d. **Trasați o histogramă reprezentând frecvența absolute (pentru toate cele 100 de valori).**

Trasarea histogramei este descrisă în secțiunea [2.2.5. Trasarea histogramei](#).

- e. Știind că toleranțele pentru diametrul exterior sunt de $10 \pm 0.3 \text{ mm}$, calculați limita superioară de toleranță (LST), limita inferioară de toleranță (LIT) și lungimea câmpului de toleranțe;
- f. Calculați limitele de control (inferioară și superioară) și media mediilor ($\bar{\bar{x}}$).

Formulele de calcul pentru cele două limite sunt:

$$LIC = \bar{x} - 3s$$

$$LIS = \bar{x} + 3s$$

Unde \bar{x} este media eșantionului iar s este abaterea standard a eșantionului.

- g. Trasați cartela de control (vezi Exemplu de raport).

Pe cartela de control trebuie să apară: LIT și LST, media mediilor ($\bar{\bar{x}}$), și cele 20 de medii calculate.

- h. Calculați capabilitatea C_p C_{ps} C_{pi} și C_{pk} .

Formulele de calculul a capabilității se găsesc în Figura 5.3. din secțiunea [5.2.1. Capabilitate](#)

- i. Calculați coeficientul de corelație Pearson între valorile primelor două eșantioane. Folosiți funcția CORREL() pentru determinarea coeficientului de corelație.
- j. Trasați linia de regresie pentru cele două seturi de date. Afișați coeficientul R^2 și ecuația liniei de regresie.
- k. Trageți o concluzie referitoare la performanța procesului în funcție de valoare C_p obținută.
- l. Generați un raport după cel din exemplu.

5.4. Resurse

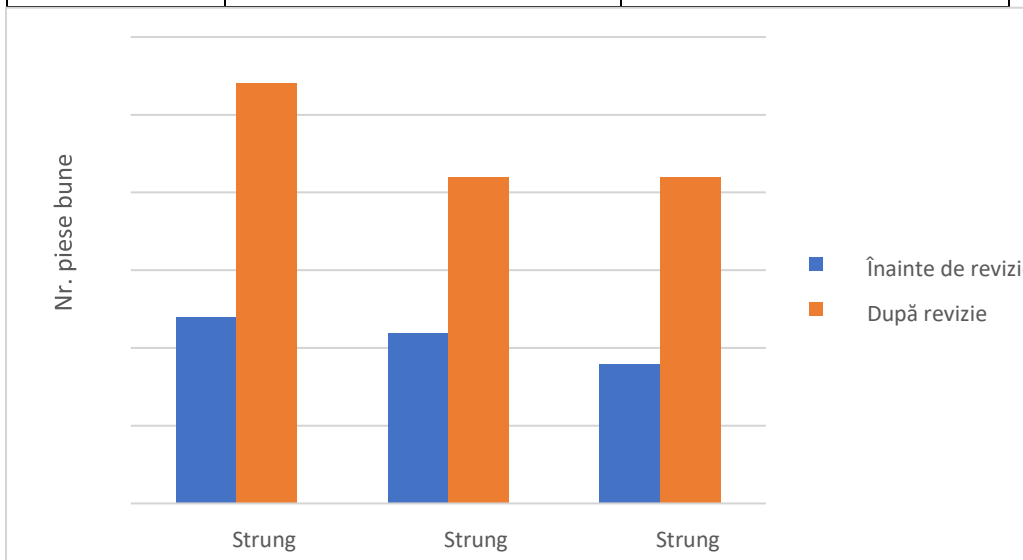
Descriere resursa	Limba	Format	Link
Definiția capabilității	EN	text	http://www.statisticalprocesscontrol.info/glossary.html
Proces capabil/stabil	EN	text	http://www.qualitymag.com/articles/93141-keys-for-maximizing-the-benefits-of-your-spc-program

Anexa 1 - Exemple de rapoarte

Exemplu raport pentru lucrarea 1

Colectarea și procesarea primară a datelor

Starea Strung	Înainte de revizie			După revizie		
	Strung 1	Strung 2	Strung 3	Strung 1	Strung 2	Strung 3
Data	17.05.2020			01.03.2021		
Operator	17.05.2020	17.05.2020		01.03.2021	01.03.2021	
Min	Georgescu	Danescu	Popescu	Georgescu	Danescu	Popescu
Max	8.927	8.744	9.131	11.075	9.677	9.616
Amplitudine	10.973	10.622		10.431	10.462	
Nr. piese bune	2.148	2.229	1.491	0.675	0.815	0.870
	12	11	9	27	21	21



Concluzie: Care a fost efectul reviziei per total? Unde a avut revizia efectul cel mai mare? Dar cel mai puțin?

Nume Prenume 1 Grupa

Exemplu raport pentru lucrarea 2

Determinarea indicatorilor statistici de localizare și împrăștiere și reprezentarea grafică a datelor

Prima linie de producție

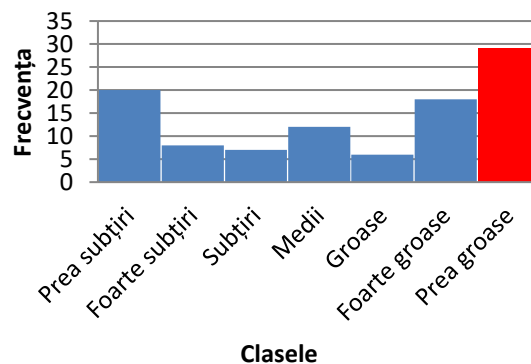
Tab. 1.1. Indicatori statistici

Indicator		Linia 1
Media aritmetică	Mx	3.6575
Media geometrică	Mg	3.6265
Media armonică	Ma	3.5946
Media pătratică	Mp	3.6873
Mediana	Me	3.7145
Valoare centrală	Xc	3.6786
Minim	Min	2.6596
Maxim	Max	4.6976
Dispersie	s ²	0.2212
Ab. Standard	s	0.4703
Amplitudine	R	2.0380

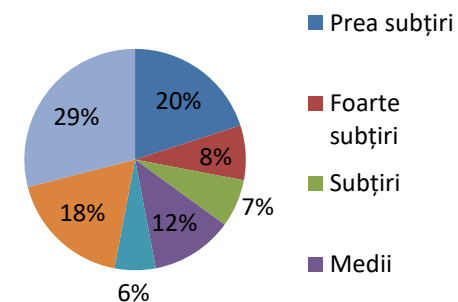
Rezultat:

Rezultatele indică o proporție de 49% rebut: 20% piese prea subțiri și 29% piese prea groase. Se recomandă o analiză a procesului pentru identificarea cauzei gradului ridicat de rebut.

Frecvența absolută a pieselor de pe prima linie de producție



Frecvența relativă a pieselor de pe prima linie de producție



Tab. 1.2. Tabelul de frecvențe

Clasa	Valori limită		Frecvența Absolută	Frecvența Relativa	Frecvența Abs. Cum. Cresc.	Frecvența Abs. Cum. Dezcresc.	Frecvența Rel. Cum. Cresc.	Frecvența Rel. Cum. Dezcresc.
Prea subțiri	2.660	3.200	20	0.20	20	100	0.20	1.00
Foarte subțiri	3.200	3.350	8	0.08	28	80	0.28	0.80
Subțiri	3.350	3.500	7	0.07	35	72	0.35	0.72
Medii	3.500	3.650	12	0.12	47	65	0.47	0.65
Groase	3.650	3.800	6	0.06	53	53	0.53	0.53
Foarte groase	3.800	3.950	18	0.18	71	47	0.71	0.47
Prea groase	3.950	4.698	29	0.29	100	29	1.00	0.29
Total	2.660	4.698	100	1.00				

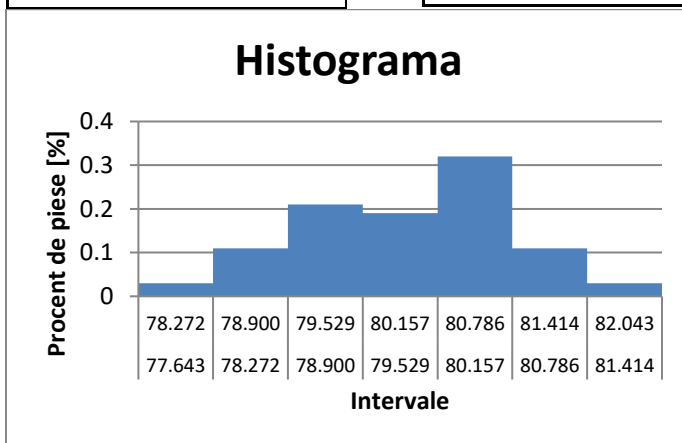
Exemplu raport pentru lucrarea 3

Eliminarea valorilor aberante și determinarea coeficientului de rebut

Cuartile	
Q1	79.242
Q2	80.036
Q3	80.492
IQR	1.250
LI1	77.367
LI2	82.367
Indicatori*	
Min	77.643
Max	82.043
Amp	4.400
Mx	79.918
s ²	0.804
s	0.896
Lungimea intervalului	
d	0.629

Intervale		ai	fi
77.643	78.272	3	0.03
78.272	78.900	11	0.11
78.900	79.529	21	0.21
79.529	80.157	19	0.19
80.157	80.786	32	0.32
80.786	81.414	11	0.11
81.414	82.043	3	0.03
Total		100	1

Calculul coeficientului de rebut			
L_inf	78	L_sup	82
x_inf	0.356792	x_sup	0.042743
x_inf%	0.567683	x_sup%	0.068008
A_inf	1.7%	A_sup	0.2%
Coef. de rebut p%			1.9%



Concluzie: Câte valori aberante au fost găsite și eliminate din calcul?

Cât este coeficientul de rebut? Din acesta, câte sunt mai mici decât trebuie și câte sunt mai mari decât trebuie?

Recomandați o îmbunătățire a procesului bazată pe rezultatele voastre

Exemplu raport pentru lucrarea 4

Estimarea parametrilor populației

Estimarea	Mediei									
	Cunoscută					Necunoscută				
Dispersia:										
Esantioane din lot nr.:	1	2	3	4		5	6	7	8	
Tipul riscului:	RUS	RUD	RBS	RBA		RUS	RUD	RBS	RBA	
Nivelul riscului	4%	5%	3%	5%	2%	10%	5%	20%	5%	2%
Media eșantionului	20.050	19.963	19.973	20.026		11.943	12.014	11.963	11.996	
Abaterea standard eșantion	0.128	0.155	0.124	0.156		0.145	0.191	0.106	0.177	
Abaterea standard populație	0.160	0.160	0.160	0.160		necunoscută				
Statistica: z , t , χ^2	1.750	1.645	1.880	1.645	2.055	1.311	1.699	1.311	1.699	2.462
Limită inf. interval încredere	19.999	-	19.918	19.978	-	11.888	-	11.930	11.954	-
Limită sup. interval încredere	-	20.011	20.028		20.086	-	12.060	11.996	-	12.038

Estimarea	Dispersiei					
	Esantioane din lot nr.:	9	10	11	12	
Tipul riscului:	RUS	RUD	RBS	RBA		
Nivelul riscului	2.5%	5%	20%	2%	5%	
Media eșantionului	35.037	35.017	35.021	35.004		
Abaterea standard eșantion	0.180	0.168	0.204	0.138		
Statistica: z , t , χ^2	16.047	42.557	19.768	39.087	15.574	42.557
Limită inf. interval încredere	0.325	-	0.151	-	0.094242	
Limită sup. interval încredere	-	0.115	0.299	0.258	-	

Raportați câte un rezultat la alegere din fiecare scenariu (estimarea mediei când dispersia este sau nu cunoscută și estimarea dispersiei) sub forma: "S-a estimat media populației (dispersia = 0.16) cu un risc unilateral stânga de 5%. Media eșantionului este 20.04 iar abaterea standard a eșantionului este 0.12.

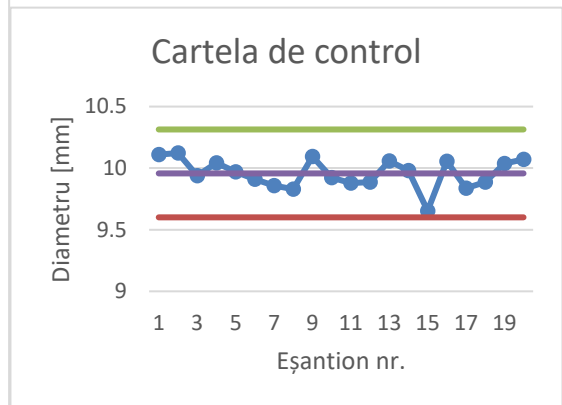
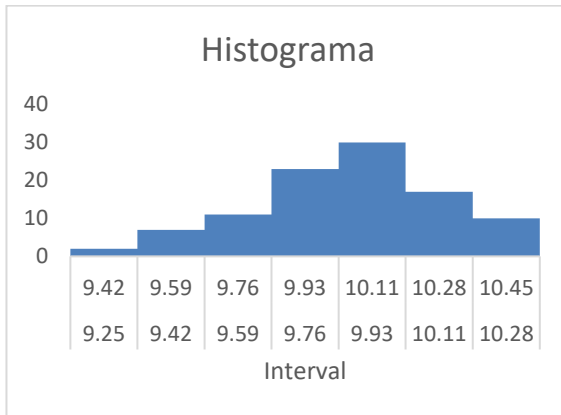
Media populației este mai mare de 19.99 cu un nivel de încredere de 95% ($M=20.04$, $SD=0.12$, $\alpha=0.05$, $z=1.645$). "

Exemplu raport pentru lucrarea 4

Calculul capabilității

Descrierea eșantionului	
x_bar	9.957
s	0.243
s^2	0.059
Min	9.253
Max	10.447
R	1.194
d	0.171
Nr. Intervale	7

Intervale		ai	fi
9.253	9.423	2	0.02
9.423	9.594	7	0.07
9.594	9.764	11	0.11
9.764	9.935	23	0.23
9.935	10.105	30	0.3
10.105	10.276	17	0.17
10.276	10.447	10	0.1
Total		100	1

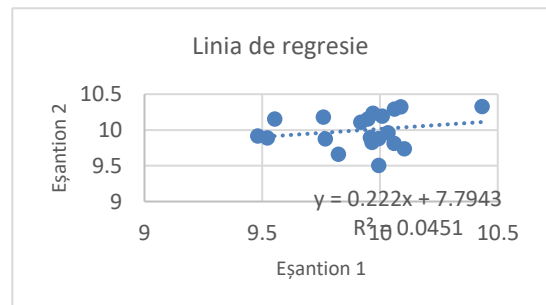


Limite de toleranță	
LIT	9.700
LST	10.300
T	0.600

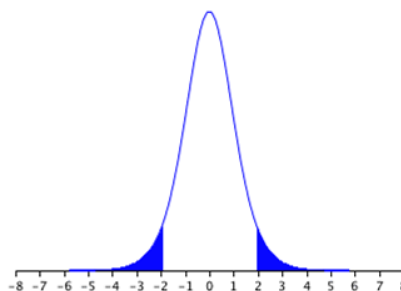
Param. cartela control	
LIC	9.601
LSC	10.314
Media mediilor	9.957

Capabilitatea proc.	
Cp	0.412
Cps	0.471
Cpi	0.353
Cpk	0.353

r (Eșant. 1 - Eșant. 2)	0.21
Concluzie	
<p>Analizați cartela de control. Aveți anomalii în proces? Explicați. Este procesul Dvs. capabil? De ce? Sunt cele două eșantioane corelate? Cât este de puternică corelația și ce sens are (pozitiv/negativ)?</p>	



Anexa 3 - Tabelul distribuției Student (valori t)

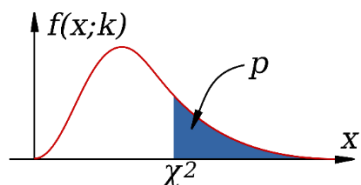


Nivelul de semnificație (α)

Grade de libertate	0.005 (1)	0.01 (1)	0.025 (1)	0.05 (1)	0.10 (1)	0.25 (1)
	0.01 (2)	0.02 (2)	0.05 (2)	0.10 (2)	0.20 (2)	0.50 (2)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.878	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
≥30	2.575	2.327	1.960	1.645	1.282	.675

* (1) - unilateral; (2) - bilateral simetric

Anexa 4 - Tabelul distribuției Chi-pătrat (valori χ^2)



Grade de libertate	Aria de la dreapta valorii critice (χ^2)							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Tabel de figuri

Fig. 1.1. Generarea numerelor aleatorii	12
Fig. 1.2. Fereastra de sortare.....	15
Fig. 1.3. Fereastra de import de date	16
Fig. 1.4. Datele importate în formatul de tabel.....	17
Fig. 1.5. Fereastra Go To Special cu Blanks selectat	18
Fig. 1.6. Fereastra de filtre (Filter).....	20
Fig. 1.7. Inserarea de grafice (Column, Line, Bar).....	21
Fig. 1.8. Reprezentarea grafică după adăugarea tuturor elementelor	22
Fig. 2.1. Distribuții unimodale, bimodale sau multimodale	35
Fig. 2.2. Exemplu de histogramă	44
Fig. 2.3. Fereastra Histogram.....	45
Fig. 2.4. Rezultatul obținut prin folosirea Histogram	45
Fig. 3.1. Fereastra de eșantionare	57
Fig. 3.2. Împărțirea unui șir în cuartile.....	58
Fig. 3.3. Determinarea grafică a coeficientului de rebut.....	61
Fig. 3.4. Exemplu de diagramă Boxplot	65
Fig. 4.1 Legătura între intervalul de încredere, nivel de încredere și risc	72
Fig. 4.2. Distribuția normală cu Risc Bilateral Simetric.....	74
Fig. 4.3. Exemplu de calcul cu distribuția normală și RBS	75
Fig. 4.4. Distribuția normală cu Risc Bilateral Asimetric.....	75
Fig. 4.5. Exemplu de calcul cu distribuția normală și RBA	77
Fig. 4.6. Distribuția normală cu Risc Unilateral Stânga.....	77
Fig. 4.7. Exemplu de calcul cu distribuția normală și RUS	78
Fig. 4.8. Distribuția normală cu Risc Unilateral Dreapta	79
Fig. 4.9. Exemplu de calcul cu distribuția normală și RUD	80
Fig. 4.10. Distribuția Student cu Risc Bilateral Simetric	81
Fig. 4.11. Exemplu de calcul cu distribuția Student și RBS	82
Fig. 4.12. Distribuția Student cu Risc Bilateral Asimetric	82
Fig. 4.13. Exemplu de calcul cu distribuția Student și RBA.....	84
Fig. 4.14. Distribuția Student cu Risc Unilateral Stânga	84
Fig. 4.15. Exemplu de calcul cu distribuția Student și RUS.....	85

Fig. 4.16. Distribuția Student cu Risc Unilateral Dreapta	86
Fig. 4.17. Exemplu de calcul cu distribuția Student și RUD	87
Fig. 4.18. Distribuția Chi-pătrat cu Risc Bilateral Simetric.....	88
Fig. 4.19. Exemplu de calcul cu distribuția Chi-pătrat și RBS.....	89
Fig. 4.20. Distribuția Chi-pătrat cu Risc Bilateral Asimetric.....	89
Fig. 4.21. Exemplu de calcul cu distribuția Chi-pătrat și RBA	91
Fig. 4.22. Distribuția Chi-pătrat cu Risc Unilateral Stânga.....	91
Fig. 4.23. Exemplu de calcul cu distribuția Chi-pătrat și RUS	92
Fig. 4.24. Distribuția Chi-pătrat cu Risc Unilateral Dreapta.....	93
Fig. 4.25. Exemplu de calcul cu distribuția Chi-pătrat și RUD.....	94
Fig. 5.1 Capacitatea proceselor de a fi controlabile și performante	104
Fig. 5.2 Elementele componente ale cartelelor de control statistic	107
Fig. 5.3 Interpretarea cartelelor de control.....	109
Fig. 5.4. Tipuri de corelații	110

Tabel de tabele

Tab. 1.1. Exemplu de tabel pentru sortarea datelor	14
Tab. 1.2. Formatele corespunzătoare fiecărei categorii de date	24
Tab. 1.3. Operatorii pe cele trei strunguri.....	25
Tab. 2.1. Mediana în cazul unui șir cu un număr impar de valori	34
Tab. 2.2. Mediana în cazul unui șir cu un nr. par de valori.....	35
Tab. 2.3. Valorile lungimilor pieselor măsurate (în mm).....	40
Tab. 2.4. Frecvențele valorilor din cele trei intervale.....	40
Tab. 2.5. Modul de calcul al frecvențelor pentru exemplul dat	42
Tab. 2.6. Exemplu de tabel pentru datele primare	47
Tab. 2.7. Valorile limită pentru clasele de calitate	48
Tab. 3.1. Tabelul frecvențelor.....	60
Tabel 5.1. Calculul capabilității proceselor și mașinilor	105