Probabilităţi şi Statistică - Curs 8

Olariu E. Florentin

Aprilie, 2017

Table of contents

- Rădăcinile cuvântului Statistică sunt latine: Status (latina veche) care înseamnă means stat (politic), Statista (italiană) înseamnă politician.
- La mijlocul secolului al XVII-lea într-o Universitate Germană a fost folosit pentru prima oară cuvântul statistik cu sensul de ştiinţă politică a statelor: analiza datelor privind statele.
- În Marea Britanie la sfârşitul secolului XVIII termenul de statistică a fost introdus cu un înţeles similar: ştiinţa statelor (sau aritmetica politică).
- Utilizarea statisticii fără a o numi în mod expres datează de la începutul civilizaţiei umane: forme incipiente de recensământ al populaţiei, sistematizarea datelor geografice şi economice etc.

- Primul studiu statistic este considerat în general a fi cel care a pus bazele demografiei: în 1662 doi englezi au introdus metode statistice cum ar fi tabelele speranţei de viaţă şi probabilităţile de supravieţuire la diferite vârste.
- Abia în secolul XIX rezultatele din teoria probabilităţilor au început a fi folosite în raţionamentul statistic.
- Bazele matematice ale statisticii s-au consolidat datorită rezultatelor profunde obţinute în teoria probabilităţilor din secolul anterior.
- Începând cu secolul XX au fost dezvoltate noi metode şi teorii, iar o influenţămajoră asupra statisticii a avut-o dezvoltarea informaticii.

"Statistics has become the universal language of the sciences."

Elementary Statistics, R. Johnson, P. Kuby

O *problemă tipică de statistică* este format din

- unul sau mai multe experimente aleatoare din efectuarea cărora rezultă o serie de date.
- o metodă de extragere a informaţiei din date si de interpretare a rezultatelor.

Modul în care informaţi aeste procesată şi interpretată dă naştere la doua ramuri ale statisticii ca ştiinţă

- *Statistica descriptivă* - colectează, prezintă şi descrie datele (de multe ori în formă grafică).
- *Statistica inferenţială* - folosind datele deja colectate ia decizii relativ la populaţia în cauză.

Definition 1

Statistica *este ştiinţa colectării, descrierii, interpretării datelor şi luării de decizii pe baza acestor date.*

Cele două ramuri ale statisticii sunt şi cei doi paşi dintr-un studiu statistic:

- *statistica descriptivă* are rolul de a
  - sintetiza, aduna şi reprezenta datele;
  - aranja informaţia, pregătind-o pentru luarea deciziilor;
- *statistica inferenţială* are drept scop
  - luarea deciziilor pe baza datelor strânse;
  - estimarea parametrilor (cum sunt media, dispersia etc);
  - verificarea ipotezelor statistice.

- Statistica îşi are propriul limbaj, dincolo de împărţirea în descriptivă şi inferenţială.
- Cel mai important concept în statistică este acela de *populaţie*: colecţia completă (exhaustivă) a obiectelor care prezintă interes pentru cel care face studiul.
- Exemple de populaţii: mulţimea studenţilor din Iaşi, mulţimea românilor analfabeţi, mulţimea dozelor de cola produse într-o luna într-o fabrică, mulţimea furtunilor tropicale din 2015.

Definition 2

O **populaţie** *este o mulţime de obiecte (numite şi indivizi) ale căror proprietăţi vor fi analizate.*

- O populaţie poate fi finită (dacă poate fi teoretic listată) sau infinită (populaţia cutremurelor de pământ din zona Vrancea).

- Din cauza dimensiunilor mari ale unei populaţii studiul statistic se concentrează asupra unei porţiuni mai mici a populaţiei. Acesta este un *eşantion* care constă din indivizi selectaţi din populaţie.

Definition 3

Un **eşantion** *este o submulţime a populaţiei. Dintr-un punct de vedere teoretic fiecare individ are aceleaşi şanse de a aparţine eşantionului, şi orice grup particular de indivizi este ales în mod independent pentru a face parte din eşantion. Dacă aceste condiţii sunt îndeplinite atunci avem un* **eşantion aleator simplu**.

- Când se alege o populaţie sau un eşantion pentru studiu, interesează o anumită trăsătură a indivizilor.
- Astfel de trăsături (atribute) pot fi: înălţimea, volumul, magnitudinea pe scara Richter, vârsta, presiunea sângelui, culoarea ochilor, suprafaţa etc.

Definition 4

O **variabilă** *sau un* **atribut** *este o caracteristică a indivizilor din populaţie sau eşantion.*

- După ce alegem un eşantion trebuie să măsurăm valorile unuia sau mai multor variabile asociate. Acestea sunt *datele*, ele pot fi numere reale, întregi, cuvinte, litere etc.

Definition 5

Datele *sunt valorile variabilei colectate de la fiecare individ din eşantion.*

- O populaţie este descrisă numeric de *parametri* (medie, disperse, deviaţie standard); parametrii sunt în centrul unui studiu statistic.

Definition 6

Un **parametru** *este o valoare numerică care priveşte întreaga populaţie.*

- Dacă populaţia este foarte mare (ceea ce se întâmplă adesea) un parametru anume nu poate fi calculat.
- O soluţie este de a calcula parametrul doar pentru un eşantion al populaţiei. Aceasta este o *statistică*.
- Pentru orice parametru şi fiecare eşantion există o statistică corespunzătoare.

Definition 7

O **statistică** *este un parametru calculat pentru un eşantion în locul întregii populaţii.*

- *Populaţie*: mulţimea studenţilor din primul an din Iaşi.
- *Eşantion*: studenţii din primul an de la FII.
- *Variabilă/atribut*: dimensiunea vocabularului lor curent.
- *Date*: 4200, 3520, 1800, ... - dimensiunile vocabularului pentru fiecare student din primul an de la FII.
- *Parametru*: media dimensiunii vocabularului studenţilor din primul an din Iaşi
- *Statistică*: media dimensiunii vocabularului studenţilor din primul an de la FII.

- Clasificarea variabilelor împarte atributele în cantitative sau calitative. Astfel există
- Variabile care oferă o *informaţie calitativă*, cum ar fi culoarea ochilor studenţilor, genurile literare ale cărţilor dintr-o bibliotecă (ficţiune, ştiinţă, literatură motivaţională etc), tipul de personalitate ale peroanelor dintr-o comunitate (sanguin, coleric, melancolic sau flegmatic), nivelul de satisfacţie a clienţilor unui magazin etc.
- Variabile care dau o *informaţie cantitativă*; spre exemplu: înălţimea studenţilor, greutatea lor, suma de bani pe care un student o cheltuie pe cărţi într-un an şcolar ş. a.

Definition 8

O **variabilă calitativă** *(sau* **categorică***) este o variabilă care descrie un individ dintr-o populaţie (conform unor categorii). O* **variabilă cantitativă** *este o variabilă care măsoară un individ dintr-o populaţie.*

- Variabilele calitative pot fi *nominale* sau *ordinale*.
- Variabilele nominale sunt: culoarea ochilor, tipul de personalitate, numele membrilor unei comunităţi etc.
- Exampe de variabile ordinale: nivelul de satisfacţie a clienţilor, nivelul de educaţie (liceal, post liceal, universitar, post universitar, doctoral) etc.

Definition 9

O **variabilă nominală** *este o variabilă care numeşte sau descrie un individ dintr-o populaţie fără a putea asigna o ordine naturală acestor valori..*

O **variabilă ordinală** *este o variabilă ale cărei valori pot fi ordonate în mod natural.*

- Variabilele cantitative pot fi *discrete* sau *continue*. Cele două tipuri pot fi distinse astfel: unele numără iar celelalte măsoară.

- O variabilă discretă de obicei numără: numărul de credite ale unui student, numărul de pagini ale unei cărţi etc; câteodată o asemenea variabilă sumează puncte/note care nu pot fi continue.
- O variabilă continuă măsoară: volumul, înălţimea, viteza, presiunea etc.

Definition 10

O **variabilă discretă** *este o variabilă care are un număr finit sau infinit dar numărabil de valori; o astfel de variabilă poate avea valori corespunzând unor puncte izolate de pe un interval real.*

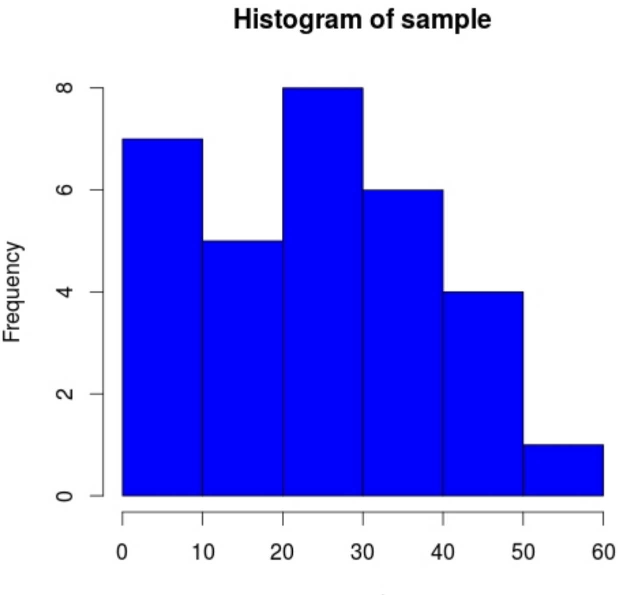
O **variabilă contină** *este o variabilă care are un număr infinit şi nenumărabil de valori; o astfel de variabilă poate avea, de obicei, orice valoare dintr-un interval real, incluzând orice valoare posibilă dintre orice două valori.*

- O primă formă de explorare a datelor este utilizarea reprezentărilor grafice care pot revela un comportament sistematic (un şablon) al variabilei.
- Tipul de reprezentare grafică depinde în mod normal de tipul variabilei.
- Pentru date calitative reprezentările grafice folosite sunt pie charts şi bar graphs.
- Pentru datele cantitative scopul reprezentărilor grafice este de a afla forma distribuţiei variabilei.

- Datele calitative sunt mai întâi transformate în frecvenţe.
- *Frecvenţa* unei observaţii (o valoare a unei variabile) este numărul de repetări ale acelei observaţii în eşantion.
- *Frecvenţa relativă* a unei observaţii este raportul dintre frecvenţa observaţiei respective şi numărul total de observaţii (dimensiunea eşantionului).
- *Distribuţia frecvenţelor* unei variabile calitative este familia tuturor perechilor formate din observaţie şi frecvenţa sa corespunzătoare.

- Pentru date cantitative putem utiliza frecvenţele şi frecvenţele relative sau *gruparea* datelor pentru a regăsi distribuţia frecvenţelo
  - datele sunt *grupate* în clase (sau *bins*) care sunt uzual intervale cu aceeaşi lungime; clasele nu trebuie sa se acopere.
  - o regulă pentru determinarea lungimii claselor: $1+\log n/\log 2$ unde n este dimensiunea eşantionului.
  - apoi datele sunt *sortate* pe clase: se determină numărul observaţiilor din fiecare clasă - acestea sunt frecvenţele.
  - suma frecvenţelor este dimensiunea eşantionului (n).
  - frecvenţele relative se pot afla împărţind frecvenţele la n.

- Cea mai utilizată metodă de reprezentare grafică a datelor cantitative este *histogram*.
- O altă metodă la îndemână pentru eşantioanele relativ mici este *stem-and-leaf*.


Histogram of sample
Frequency
0
2
4
6
8
0
10
20
30
40
50
60

- Când privim reprezentarea grafică a datelor din eşantion ne putem pune următoarele întrebări.
- Care sunt valorile centrale/medii?
- Cât de mult sunt împrăştiate aceste date în jurul valorilor medii?
- Care este froma distribuţiei?
- Există valori care nu se potrivesc cu imaginea generală a distribuţiei?

- *Tendinţa centrală* sau *centrul distribuţiei* este centrul (abstract) al datelor. Toate măsurile tendinţei centrale sunt legate într-un fel sau altul de noţiunea de medie.
- Diferite moduri de a defini tendinţa centrală:
  - Punctul care ţine distribuţia în echilibru.
  - Numărul care minimizează suma tuturor deviaţiilor absolute.
  - Numărul care minimizează suma tuturor deviaţiilor la pătrat.
  - Cea mai frecventă valoare.

- Să presupunem că valorile din eşantion sunt $x_1, x_2, \ldots, x_n$.

Definition 11

Media de selecţie *sau* **media eşantionului** *este media aritmetică a tuturor datelor din eşantion:*

$$\overline{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

- Formula mediei pentru întreaga populaţie este în esenţă identică.
- *Media populaţiei* se notează cu μ.

- În limbajul teoriei probabilităţilor media populaţiei este media unei variabile aleatoare, X, ale cărei valori sunt sunt cele ale indivizilor din populaţie; deci $M[X] = \mu$.
- Media de selecţie este o statistică care estimează media populaţiei.
- Să presupunem că $X_1, X_2, \dots, X_n$ sunt variabilele din spatele fiecărui individ al eşantionului, iar x_i este doar o valoare a variabilei X_i.
- Atunci X_i este o variabilă aleatoare cu aceeaşi distribuţie ca a a lui X. Mai mult, variablele $(X_i)_{1 \leqslant i \leqslant n}$ sunt independente în ansamblu.
- Aceste observaţii conduc la faptul ca media de selecţie poate fi văzută ca o variabilă aleatoare, iar media aritmetic calculată pentru un eşantion este una dintre posibilele valori ale ei (fiecare eşantion dă o altă valoare a mediei de selecţie).

- Dacă media de selecţie este o variabilă aleatoare, îi putem calcula media:

$$M[\overline{x}_n] = M\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] =$$

$$= \frac{1}{n} M\left[X_1 + X_2 + \cdots + X_n\right] =$$

$$= \frac{M[X_1] + M[X_2] + \cdots + m[X_n]}{n} = \mu.$$

- Media mediei de selecţie este media populaţiei.
- O astfel de statistică se numeşte *estimator nedeplasat* al parametrului corespunzător.
- Media de selecţie este un estimator nedeplasat pentru media populaţiei.

- Formula din definiţia anterioară este valabilă pentru date negrupate. În acest caz toate datele din eşantion contribue direct la calulul mediei de selecţie.
- Pentru date grupate se foloseşte o formulă cu ponderi:

$$M = \frac{m_i * f_i}{\sum_i f_i}$$

unde m_i este mijlocul intervalului clasei i, iar f_i este numărul de observaţii care aparţin clasei i.

- În această formulă observaţiile nu contribuie direct la calculul mediei; cu toate acestea este o formulă preferată în cazul datelor grupate pentru eşantioane mari fiind mai uşor de calculat.

- Ne întoarcem acum la definiţia iniţială (pentru date negrupate) a mediei de selecţie.
- Variaţii mici în suma de la numărător nu modifică prea mult media. Spunem ca media este stabilă la variaţii mici ale datelor.
- *Valorile aberante* sau *extreme* pot avea o influenţă mare asupra mediei; introducând o valoare foarte mare sau foarte mică media se poate schimba foarte mult.
- Media este o funcţie *liniară* (la fel ca media unei variabile aleatoare).
- *Deviaţiile* de la medie sunt $(x_i - \overline{x}_n)$; suma lor este zero:

$$\sum_i (x_i - \overline{x}_n) = 0.$$

- (*Definiţia variaţională*) Se poate arăta că media este numărul M care minimizează suma deviaţiilor la pătrat:

$$\sum_i (x_i - M)^2.$$

- Există şi alte tipuri de medie în afară de cea artimetică (A): media geometrică (G), media armonică (H).

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n},\ H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}}.$$

- Să presupunem că o maşină parcurge distanţa dintre două oraşe de patru ori cu vitezele 80km/h, 90km/h, 60km/h, and 120km/h, respectiv. Care a fost viteza sa medie?
- Folosim media aritmetică obţinem 87.5 km/h; dar media adecvată aici este cea armonică: 82.3km/h.

- Mediana este o *statistică ordonată*; calculul unei astfel de statistici presupune ordonarea crescătoare a datelor din eşantion.

Definition 12

Median *(Me) este valoarea din mijloc când datele din eşantion sunt sortate.*

- Mediana împarte datele din eşantion în două jumătăţi: o jumătate conţine datele mai mari sau egale decât mediana, iar cealaltă jumătate le conţine pe cele mai mici sau egale.
- Valoarea medianei este o observaţie sau media a doua observaţii (pentru eşantioane de dimensiune pară).
- Ca statistică mediana este mult mai puţin influenţată de existenţa valorilor aberante.

Definition 13

Mòdul *este observaţia cea mai frecventă din eşantion.*

- Pentru date grupate se alege mai întâi clasa cu cea mai mare frecvenţă *clasa modală*. Fie i indexul acestei clase, a_i marginea stângă a intervalului corespunzător şi L lungimea comună a intervalelor.
- Atunci mòdul poate fi calculat folosind formula

$$mod = a_i + \frac{L * (f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_i - f_{i+1})}.$$

- *Antimòdul* este cea mai puţin frecventă observaţie.

- Mai stabile la valorile aberante sunt mediana şi mòdul.
- Media incorporează toate valorile şi nu poate fi calculată, în cazul datelor grupate pentru distribuţii deschise (primul, sau ultimul interval deschis).
- Mediana şi mòdul nu sunt funcţii liniare.
- Mòdul este calculat mai laes pentru date grupate.
- Pentru distribuţii asimetrice mòdul oferă cea mai reală imagine asupra tendinţei centrale.

- Dacă eşantionul conţine date foarte mari sau foarte mici mediana este măsura preferată mediei - stabilitatea o face mai reprezentativă.
- Pentru distribuţii simetrice cele trei măsuri sunt aproape egale.
- Forma distribuţiei poate fi legată ade relaţia dintre medie şi mediană; forma poate fi
- asimetrică spre stânga dacă $\overline{x}_n < Me$;
- simetrică dacă $\overline{x}_n = Me$;
- asimetrică spre dreapta dacă $\overline{x}_n > Me$;

- Relativ la măsurile tendinţei centrale există *măsuri de poziţie* care sunt statistici ordonate ca şi mediana.

Definition 14
Cvartilele *sunt valori care împart domeniul (ordonat al) observaţiilor în patru segmente egale.*

- Prima cvartilă, Q_1, este o valoare astfel în cât 25% dintre observaţii sunt cel mult egale cu Q_1 şi cel mult 75% sunt mai mari sau egale.
- A treia cvartilă, Q_3,este o valoare astfel în cât 75% dintre observaţii sunt cel mult egale cu Q_3 şi cel mult 25% sunt mai mari sau egale.

- A doua cvartilă, Q_2,este o valoare astfel în cât 50% dintre observaţii sunt cel mult egale cu Q_2 şi cel mult 50% sunt mai mari sau egale. Din acest motiv a doua cvartilă este egală cu mediana: $Me = Q_2$.
- Cvartilele au proprietăţi similare cu cele ale medianei. Cea mai importantă fiind aceea că sunt stabile în prezenţa valorilor aberante.
- Statistici ordonate similare sunt: *decilele*, *percentilele* etc. Toate aceste statistici împart datele ordonate în subeşantioane egale.
- De exemplu există nouă decile care împart datele sortate în zece părţi egale, fiecare parte reprezentând 10% din eşantion.

- După determinarea "centrului" datelor studiul statistic continuă cu analiza *împrăştierii* sau a *variabilităţii* datelor
- Valorile din eşantion pot să difere mult între ele şi faţă de valoarea "centrală".
- Măsura în care valoare "centrală"/medie este reprezentativă pentru întreg eşantionul depinde de variabilitatea (sau dispersia) datelor.
- Eşantionul are variabilitate mare dacă există valori foarte mari sau foarte mici faţă de valoarea medie.
- Deoarece avem două moduri importante de a măsura tendinţa centrală (media şi mediana) vom avea două metode de a măsura împrăştierea.

Definition 15

Domeniul *este diferenţa dintre cea mai mică şi cea mai mare valoare din eşantion.*

$$range = max - min.$$

- Deoarece definiţia aceasta se bazează doar pe valorile extreme, dacă minimul sau maximul este foarte mare respectiv foarte mic, domeniul nu este reprezentativ pentru variabilitatea datelor.
- Se observă că valorile aberante au o influenţă directă asupra domeniului.

Începem cu măsurile variabilităţii legate de medie.

- Deviaţiile faţă de medie sunt $(x_i - \overline{x}_n)$.
- O deviaţie $(x_i - \overline{x}_n)$ este pozitivă (negativă) când x_i este mai mare (mai mică) decât media de selecţie.
- Pentru a descrie o valoare medie a deviaţiilor s-ar putea utiliza media aritmetică a acestor deviaţii. Dar suma acestor deviaţii fiind zero, o astfel de medie este nulă.
- Putem îndepărta acest efect ridicând la pătrat deviaţiile şi utilizând o medie pătratică în locul uneia aritmetice.

Definition 16

Dispersia eşantionului, *s^2, n fiind dimensiunea eşantionului, este:*

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}{n-1},$$

- Dispersia eşantionului este nenegativă şi este zero dacăşi numai dacă valorile sunt toate identice.
- Dispersia eşantionului este statistica asociată *dispersiei populaţiei*, notată cu σ^2.

- Morivul pentru care se utilizează $(n-1)$ ca numitor în definiţia dispersiei eşantionului este acela că astfel se obţine un estimator nedeplasat.
- Media dispersiei eşantionului (văzută ca o variabilă aleatoare) este

$$M[s^2] = M\left[\frac{\sum_{i=1}^{n}(X_i - \overline{x}_n)^2}{n-1}\right] = M\left[\frac{\sum_{i=1}^{n}\left(nX_i - \sum_{j=1}^{n} X_j\right)^2}{n(n-1)}\right] =$$

$$= \frac{\sum_{i=1}^{n} M\left[n^2X_i^2 - 2nX_i\left(\sum_{j=1}^{n} X_j\right) + \left(\sum_{j=1}^{n} X_j\right)^2\right]}{n^2(n-1)} =$$

$$= \frac{\sum_{i=1}^{n} M\left[n^2 X_i^2\right] - 2n \sum_{i=1}^{n} M\left[\sum_{j=1}^{n} X_i X_j\right] + nM\left[\sum_{j=1}^{n} X_j^2 + 2\sum_{i<j} X_i X_j\right]}{n^2(n-1)}$$

$$= \frac{n^2 \sum_{i=1}^{n} M\left[X_i^2\right] - 2n \sum_{i=1}^{n} M\left[X_i^2\right] - 2n \sum_{i=1}^{n} \sum_{j \neq i}^{n} M\left[X_i X_j\right]}{n^2(n-1)} +$$

$$+ \frac{n \sum_{j=1}^{n} M\left[X_j^2\right] + 2n \sum_{i<j} M\left[X_i X_j\right]}{n^2(n-1)} =$$

$$= \frac{n(n-1)\sum_{i=1}^{n} M\left[X_i^2\right] - 2n\sum_{i<j} M\left[X_i X_j\right]}{n^2(n-1)} =$$

$$= \frac{n^2(n-1)\sum_{i=1}^{n} M\left[X_i^2\right] - 2n\sum_{i<j} M\left[X_i\right] M\left[X_j\right]}{n^2(n-1)} =$$

$$= \frac{\sum_{i=1}^{n} M[X_i^2] - M\left[X_i\right]^2}{n} = \sigma^2.$$

- O formulă mai simplă (exerciţiu) pentru dispersia eşantionului este

$$s^2 = \frac{n \sum_{i=1}^{n} x_i - \left(\sum_{i=1}^{n} x_i \right)^2}{n(n-1)}.$$

Definition 17

Deviaţia standard a eşantionului, *s, este rădăcina pătrată a dispersiei eşantionului.*

- Deviaţia standard a eşantionului este un estimator deplasat al deviaţiei standard a populaţiei, σ.
- Se poate arăta că media deviaţiei standard a eşantionului este mai mică decât cea a populaţiei, $M[s] < \sigma$.

Continuăm cu măsuri ale împrăştierii legate de mediană. Mai întâi sumarul celor cinci numere.

Definition 18

Sumarul celor cinci numere *este compus din*

1. *min, cea mai mică valoare din eşantion;*
2. Q_1*, prima cvartilă;*
3. Me*, mediana;*
4. Q_3*, a treia cvartilă;*
5. *max, cea mai mare valoare din eşantion.*

- O metodă grafică de a reprezenta sumarul celor cinci numere: *box-and-whiskers*.

Definition 19

Cvartila medie *este valoarea de mijloc dintre prima şi cea de-a treia cvartilă:*

$$midq = \frac{Q_1 + Q_3}{2}.$$

Domeniu intercvartilic *este diferenţa dintre prima şi cea de-a treia cvartilă:*

$$IQR = Q_3 - Q_1.$$

- *Valorile aberante* sunt acele valori din eşantion care pot fi considerate prea mici sau prea mari faţă de "tabloul" general al eşantionului.
- Evident, valorile aberante sunt legate de variabilitatea datelor. În mod obinuit aceste valori vin din erori de măsură, dar pot avea şi cauze naturale.
- Câteodată aceste valori aberante (dacă sunt datorate măsurilor) pot fi eliminate din eşantion înainte de orice altă analiză statistică.
- Vom avea două reguli de determinare a valorilor aberante, deoarece şi variabilitatea datelor se măsoară în două feluri.

- Prima regulă este legată de medie. Pot fi considerate valori aberante acele valori ale eşantionului care nu aparţin intervalului $(\overline{x}_n - 2s, \overline{x}_n + 2s)$.
- A doua regulă se numeşte regula $1.5 * IQR$ şi spune că o valoare este aberantă dacă nu aparţine intervalului $(Q_1 - 1.5 * IQR, Q_3 + 1.5 * IQR)$.

Sfârşit

- Freedman, D., R. Pisani, R. Purves, *Statistics*, W. W. Norton & Company, 4th edition, 2007.
- Johnson, R., P. Kuby, *Elementary Statistics*, Brooks/Cole, Cengage Learning, 11th edition, 2012.
- Shao, J., *Mathematical Statistics*, Springer Verlag, 1998.
- Spiegel, M. R., L. J. Stephens, *Theory and Problems of Statistics*, Schaum's Outline Series, McGraw Hill, 3rd edition, 1999.