

A3. ANALIZA VARIABILITATII GLOBALE PENTRU VARIABLE NUMERICE

OBIECTIVE:

- **cea mai probabila valoare** a variabilei numerice (continut de NH4);
- **eroarea de estimare** a celei mai probabile valori a variabilei numerice (continut de NH4).

DATE NECESARE: valorile variabilei numerice

Pregatirea fisierului cu date:

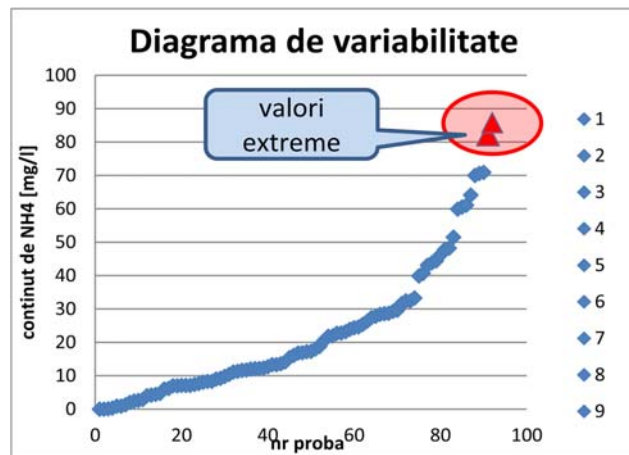
- crearea unei noi foi de lucru ("AVG_NH4") in fisierul "**Date_prelucrate.xls**";
- copierea valorilor continutului de NH4 din baza de date ("**Date_primare.xls**"), foaia "**Date**", in foaia "**AVG_NH4**".

Metodologia AVG pentru variabile numerice:

a. identificarea si analiza valorilor extreme (valori ffff mici sau ffff mari in raport cu majoritatea valorilor) prin intermediul

DIAGRAMEI DE VARIABILITATE

- *Insert – Charts – XY Scatter* + sortarea valorilor pentru o mai buna vizualizare;
- au fost identificate doua valori extreme de



maxim (82,2; 85,8 mg/l) ce vor fi eliminate din prelucrarile ulterioare.

b. calculul frecventelor absolute experimentale n_i ale continuturilor de NH4:

- estimarea **marimii intervalului de grupare** (Δ):

$$\Delta = \frac{A}{1 + 3,322 \cdot \log(N)}$$

- calculul **numarului de intervale** de grupare:

$$k = \frac{A}{\Delta}$$

- stabilirea **limitelor inferioare si superioare** ale intervalelor de grupare;
- numararea continuturilor de NH₄ din fiecare interval de grupare (**frecventele absolute experimentale** ale continuturilor de NH₄);
- verificarea corectitudinii calculului frecventelor absolute:

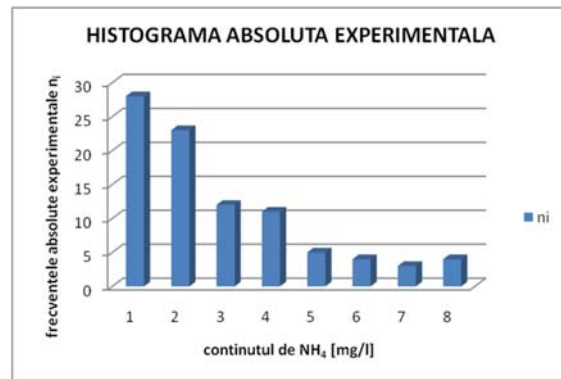
$$n_1 + n_2 + \dots + n_k = N$$

c. **reprezentarea grafica** a distributiei frecventelor absolute experimentale ale continuturilor de NH₄ – **HISTOGRAMA ABSOLUTA EXPERIMENTALA**

- *Insert – Charts – Column 2D sau 3D*

d. **evaluarea gradului de asimetrie** a histogramei absolute experimentale:

- **calitativ** – histograma absoluta este *asimetrica pozitiva* sau “de stanga” cu *intensitatea asimetriei* – SEVERA;



- **cantitativ**, utilizand coeficientul de asimetrie (**SKEWNESS**; =SKEW(...)):

- = 0 – simetrie perfecta
- > 0 – asimetrie pozitiva sau “de stanga”
- < 0 – asimetrie negativa sau “de dreapta”

- **Concluzie**: histograma absoluta experimentală a continuturilor de NH₄ prezinta asimetrie pozitiva (coef. asim. = 1,17); in acest caz, **valoarea cea mai probabila** a continuturilor de NH₄ este **SUPRAESTIMATA** de valoarea medie a continuturilor de NH₄.

e. **testarea concordantei** intre repartitia **frecventelor absolute experimentale** n_i si **modelul repartitiei normale** – **TESTUL HI^2**

Fundamentul testarii il reprezinta **riscul asumat** α. Alegem α = 10%.

- **calculul frecventelor absolute teoretice n_{ti}** (corespunzatoare modelului repartitiei normale);

$$n_{ti} = \frac{N \cdot \Delta}{s \cdot \sqrt{2 \cdot \pi}} \cdot EXP \left(-\frac{1}{2} \cdot \frac{(v_{ci} - \bar{v})^2}{s^2} \right)$$

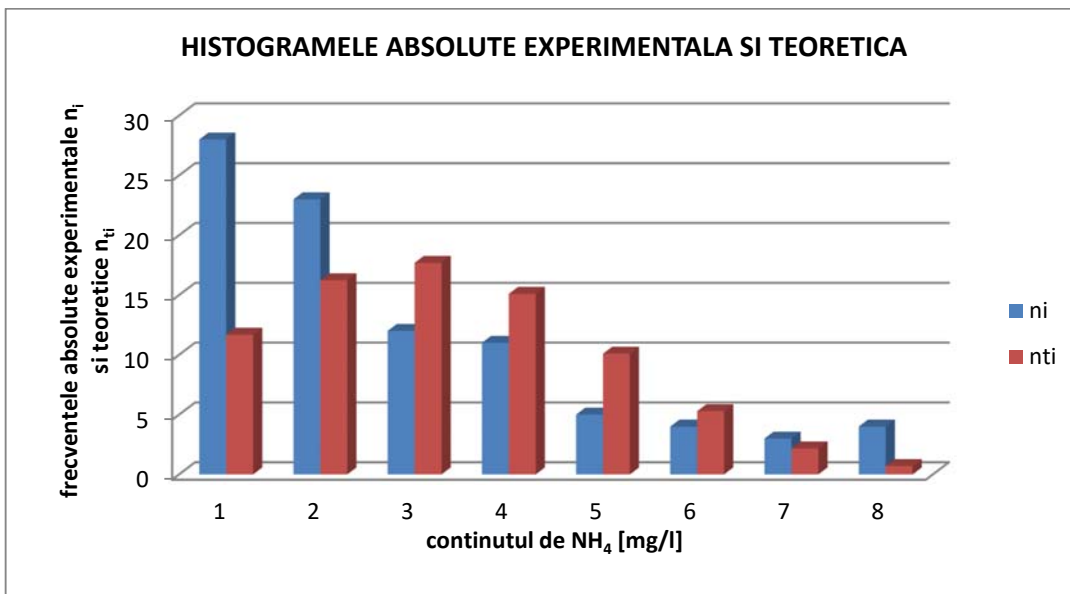
$i = 1, 2, \dots, k$

s – abaterea standard (standard deviation; =STDEV(...));

v_{ci} – valoarea centrala a fiecarui interval k de grupare;

\bar{v} - valoarea medie a continuturilor de NH_4 .

- **reprezentarea grafica** comparativa a frecventelor absolute experimentale si teoretice;



- **calculul statisticii HI^2 experimental**

$$\chi_{exp}^2 = \sum_{i=1}^{i=k} \frac{(n_i - n_{ti})^2}{n_{ti}}$$

- **calculul statisticii HI^2 teoretic** (in functie de riscul asumat α si numarul de grade de libertate ν) (=CHIINV(α , ν))

$$\nu = k - (m + 1) = k - 3$$

m – numarul de parametri statistici utilizati in calculul frecventelor absolute teoretice n_{ti} (valoarea medie si abaterea standard)

- **compararea statisticilor HI^2 experimental si HI^2 teoretic**
 - **daca $HI^2_{exp} < HI^2_{teor/adm}$** , repartitia frecventelor experimentale ale continuturilor de NH_4 **este NORMALA**.

Valoarea cea mai probabila a continuturilor de NH4 este valoarea medie.

- **daca $HI^2 \text{ exp} > HI^2 \text{ teor/adm}$** , repartitia frecventelor experimentale ale continuturilor de NH4 **NU este NORMALA**. Valoarea cea mai probabila a continuturilor de NH4 NU este valoarea medie si este SUPRAESTIMATA de catre aceasta. Se va incerca **NORMALIZAREA** distributiei frecventelor continuturilor de NH4.
- **CONCLUZIE**: intrucat **$HI^2 \text{ exp} (49,56) > HI^2 \text{ teor/adm} (9,24)$** , repartitia frecventelor absolute experimentale ale continuturilor de NH4 **NU este NORMALA**.

f. evaluarea celei mai probabile valori in cazul unei selectii de date cu repartitie "ANORMALA"

- **normalizarea distributiei frecventelor experimentale** utilizand transformari (prin intermediul unor functii analitice) aplicate continuturilor de NH4
 - **aplicarea transformarilor** valorilor originale ale continutului de NH4;
 - $T_1 = \frac{1}{v_i}$
 - $T_2 = \ln(v_i)$
 - $T_3 = \sqrt{v_i}$
 - **calculul coeficientului de asimetrie** pentru valorile transformate prin intermediul celor 3 functii analitice;
 - $\beta_{3T_1} = 5,48$
 - $\beta_{3T_2} = -1,55$
 - $\beta_{3T_3} = 0,36$
 - alegem **functia radical** (T_3) pentru normalizarea frecventelor continuturilor de NH4
 - aplicand metodologia A.V.G. si in urma aplicarii testului HI^2 , am constatat ca **distributia frecventelor valorilor transformate este NORMALA**. Astfel, valoarea cea mai probabila a valorilor transformate este **valoarea lor medie** (4,24).

$$\bar{v}_{(NH_4)T_3} = 4,24$$

- calculul **valorii celei mai probabile a continutului de NH4** se face aplicand transformarea inversa valorii medii obtinute.

$$\bar{v}_{NH_4} = 17,94 \text{ mg/l}$$

g. calculul erorii de estimare a celei mai probabile valori in cazul unei selectii de date cu repartitie "ANORMALA"

- eroarea de estimare (=CONFIDENCE(α, s, N)) depinde de:
 - riscul asumat (α);
 - numarul punctelor de observatie (N);
 - variabilitatea selectiei de date (s);

$$\varepsilon(\alpha, s, N)_{T_3} = 0,36$$

- **eroarea care afecteaza cea mai probabila valoare a continutului de NH4** se calculeaza aplicand transformarea inversa valorii $\varepsilon(\alpha, s, N)_{T_3}$ calculate mai sus:

$$\varepsilon(\alpha)_{NH_4} = 0,13 \text{ mg/l}$$

CONCLUZIA A.V.G. pentru variabile numerice (continut de NH4): cu o probabilitate de 90%, cea mai probabila valoare a continutului de NH4 este cuprinsa in intervalul [17,81 ÷ 18,07] mg/l:

- 17,94 – 0,13 mg/l – limita inferioara a intervalului – 17,81 mg/l;
- 17,94 + 0,13 mg/l – limita superioara a intervalului – 18,07 mg/l.