

### A3. ANALIZA VARIABILITATII GLOBALE PENTRU VARIABLE NUMERICE

#### OBIECTIVE:

- **cea mai probabila valoare** a variabilei numerice (continut de NH<sub>4</sub>);
- **eroarea de estimare** a celei mai probabile valori a variabilei numerice (continut de NH<sub>4</sub>).

**DATE NECESARE:** *valorile variabilei numerice*

Pregatirea fisierului cu date:

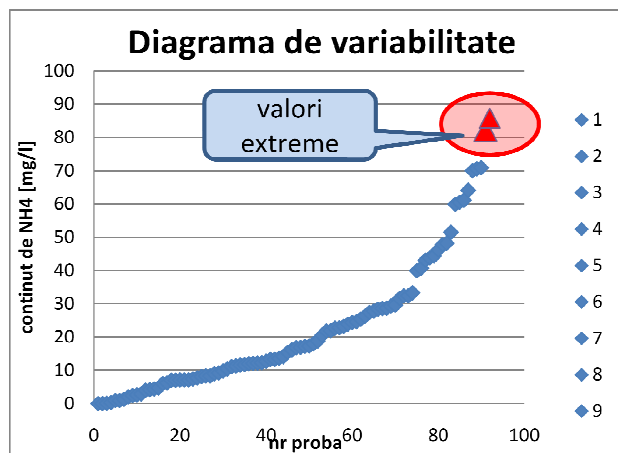
- crearea unei noi foi de lucru ("AVG\_NH<sub>4</sub>") in fisierul "**Date\_prelucrate.xls**";
- copierea valorilor continutului de NH<sub>4</sub> din baza de date ("**Baza\_de\_date.xls**", foaia "Date"), in foaia "AVG\_NH<sub>4</sub>".

**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 – 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 NH<sub>4</sub>:

- **calculul amplitudinii** continuturilor de NH<sub>4</sub> (**A**):

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

- estimarea **marimii intervalului de grupare** ( $\Delta$ ):

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

- calculul **numarului de intervale** de grupare ( $k$ ) a continuturilor de NH<sub>4</sub>:

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

- stabilirea **limitelor (inferioara si superioara)** fiecarui interval de grupare a valorilor;

$$lim_{sup} = lim_{inf} + \Delta$$

- numararea continuturilor de NH<sub>4</sub> din fiecare interval de grupare (**frecventele absolute experimentale** ale continuturilor de NH<sub>4</sub>);
- **verificarea corectitudinii calculului** frecventelor absolute experimentale:

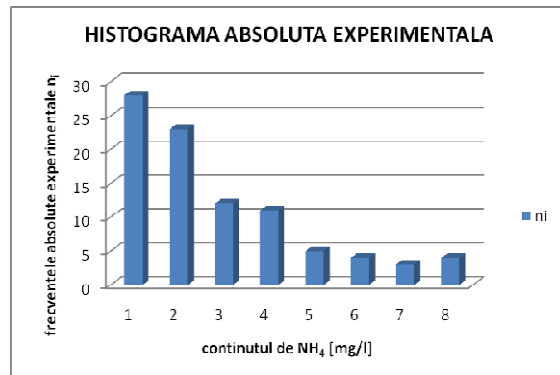
$$n_1 + n_2 + \dots + n_k = N$$

**c. reprezentarea grafica** a distributiei frecventelor absolute experimentale ale continuturilor de NH<sub>4</sub> – **HISTOGRAMA ABSOLUTA EXPERIMENTALA**

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

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

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



- **cantitativ**, utilizand coeficientul de asimetrie (**SKEWNESS**; =SKEW(v<sub>i</sub>)):

- $\beta_3 = 0$  – simetrie perfecta
- $\beta_3 > 0$  – asimetrie pozitiva sau “de stanga”
- $\beta_3 < 0$  – asimetrie negativa sau “de dreapta”

- **Concluzie**: histograma absoluta experimentală a continuturilor de NH<sub>4</sub> prezinta asimetrie pozitiva ( $\beta_3 = +1,17$ ); in acest caz, **valoarea cea mai probabila** a continuturilor de NH<sub>4</sub> este **SUPRAESTIMATA** de valoarea medie a continuturilor de NH<sub>4</sub>.

**e. testarea concordantei** între repartiția **frecvențelor absolute experimentale**  $n_i$  și **modelul repartiției normale – TESTUL  $H_1^2$**

**Fundamentul** testării îl reprezintă **riscul asumat**  $\alpha$ . Alegem  $\alpha = 10\%$ .

- **calculul frecvențelor absolute teoretice**  $n_{i\text{ teor}}$  (corespunzătoare modelului repartiției normale);

$$n_{i\text{ teor}} = \frac{N \cdot \Delta}{s \cdot \sqrt{2 \cdot \pi}} \cdot \text{EXP} \left( -\frac{1}{2} \cdot \frac{(v_{ci} - \bar{v})^2}{s^2} \right); i = 1, 2, \dots, k$$

$s$  – abaterea standard (standard deviation; =STDEV( $v_i$ ));

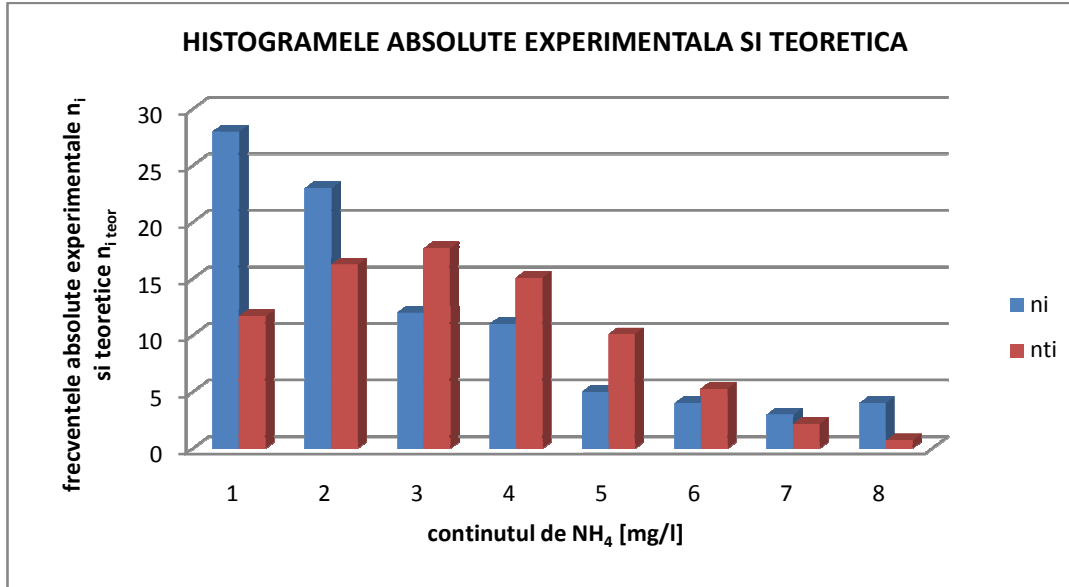
$v_{ci}$  – valoarea centrală a fiecărui interval  $k$  de grupare;

$\bar{v}$  - valoarea medie a conținuturilor de  $\text{NH}_4$  (=AVERAGE( $v_i$ )).

- **verificarea corectitudinii calculului frecvențelor teoretice**

$$n_{1\text{ teor}} + n_{2\text{ teor}} + \dots + n_{k\text{ teor}} \cong N$$

- **reprezentarea grafică** comparativă a frecvențelor absolute experimentale și teoretice;



- **calculul statisticii  $H_1^2$  experimental**

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

- **calculul statisticii  $H^2$  teoretic/admisibil** ( $=\text{CHIINV}(\alpha, \nu)$ ) se face in functie de riscul asumat ( $\alpha$ ) si numarul de grade de libertate ( $\nu$ )

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

$m$  – numarul de parametri statistici utilizati in calculul frecventelor absolute teoretice  $n_{i \text{ teor}}$  (valoarea medie si abaterea standard)

- **compararea statisticilor  $H^2$  experimental si  $H^2$  teoretic**
  - **daca  $H^2 \text{ exp} < H^2 \text{ teor/adm}$** , repartitia frecventelor experimentale ale continuturilor de  $\text{NH}_4$  **este NORMALA** pentru **riscul asumat  $\alpha = 10\%$** . Valoarea cea mai probabila a continuturilor de  $\text{NH}_4$  este valoarea medie.
  - **daca  $H^2 \text{ exp} > H^2 \text{ teor/adm}$** , repartitia frecventelor experimentale ale continuturilor de  $\text{NH}_4$  **NU este NORMALA** pentru **riscul asumat  $\alpha = 10\%$** . Valoarea cea mai probabila a continuturilor de  $\text{NH}_4$  **NU este valoarea medie** (20,99 mg/l) si este **SUPRAESTIMATA** de catre aceasta. Se va incerca **NORMALIZAREA** distributiei frecventelor continuturilor de  $\text{NH}_4$ .
  - **CONCLUZIE:** intrucat  **$H^2 \text{ exp} (49,56) > H^2 \text{ teor/adm} (9,24)$** , **repartitia frecventelor absolute experimentale ale continuturilor de  $\text{NH}_4$  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  $\text{NH}_4$ 
  - **aplicarea transformarilor** valorilor originale ale continutului de  $\text{NH}_4$ ;
    - $T_1 = \frac{1}{v_i}$
    - $T_2 = \ln(v_i)$
    - $T_3 = \sqrt{v_i}$
  - **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 NH<sub>4</sub>
- **Concluzie:** aplicand metodologia A.V.G. si in urma aplicarii testului HI<sup>2</sup>, am constatat ca **distributia frecventelor valorilor transformate cu functia radical este NORMALA** pentru **riscul asumat  $\alpha = 10\%$** . 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 NH<sub>4</sub>** se face aplicand transformarea inversa (ridicarea la patrat) valorii medii obtinute.

$$\bar{v}_{NH_4} = 4,24^2 = 17,94 \text{ mg/l}$$

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

- eroarea de estimare (=CONFIDENCE( $\alpha, s, N$ )) depinde de:
  - riscul asumat ( $\alpha$ );
  - variabilitatea selectiei de date ( $s$ );
  - numarul punctelor de observatie ( $N$ );
$$\varepsilon(\alpha, s, N)_{T_3} = 0,36$$
- **eroarea care afecteaza cea mai probabila valoare a continutului de NH<sub>4</sub>** se calculeaza aplicand transformarea inversa (ridicarea la patrat) valorii  $\varepsilon(\alpha, s, N)_{T_3}$  calculate mai sus:

$$\varepsilon(\alpha)_{NH_4} = 0,36^2 = 0,13 \text{ mg/l}$$

**CONCLUZIA A.V.G. pentru variabile numerice (continut de NH<sub>4</sub>):** cu o probabilitate de 90%, cea mai probabila valoare a continutului de NH<sub>4</sub> 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.