

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 (continut de NH4)

Pregatirea fisierului cu date:

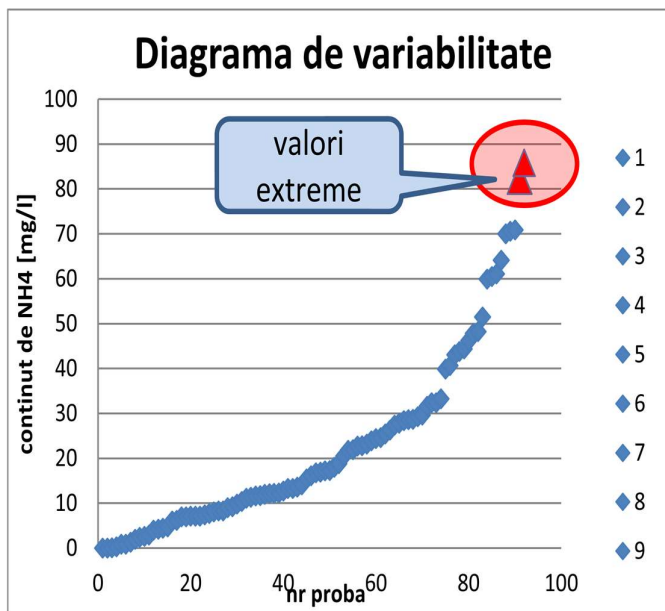
- crearea unei noi foi de lucru ("AVG_NH4") in fisierul "**Date_prelucrate.xls**";
- copierea valorilor continutului de NH4 din baza de date ("**Baza_de_date.xls**", foaia "Date"), in foaia "AVG_NH4".

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 – 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:**

- **calculul amplitudinii** continuturilor de NH4 (**A**):

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

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

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

- calculul **numarului de intervale** de grupare (k) a continuturilor de NH₄:

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

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

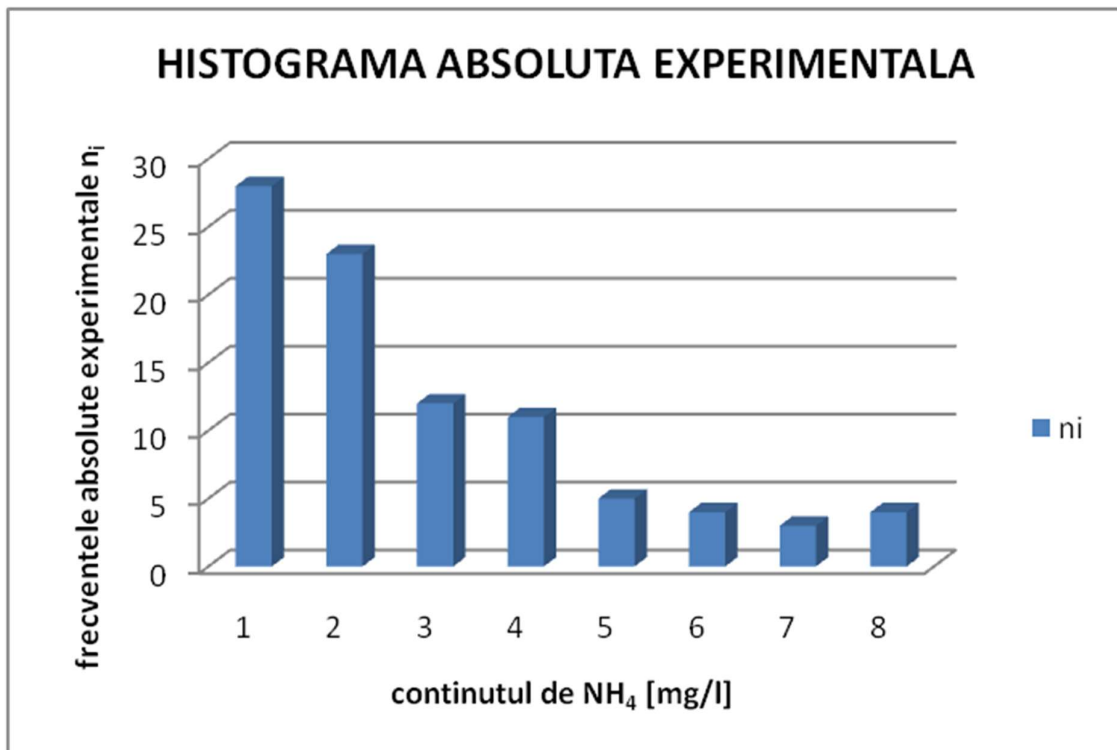
$$\lim_{sup} = \lim_{inf} + \Delta$$

- numararea continuturilor de NH₄ din fiecare interval de grupare (**frecventele absolute experimentale** $n_{i \text{ exp}}$ ale continuturilor de NH₄);
- **verificarea corectitudinii calculului** frecventelor absolute experimentale:

$$n_{1 \text{ exp}} + n_{2 \text{ exp}} + \dots + n_{k \text{ exp}} = N$$

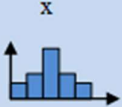
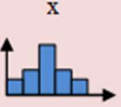
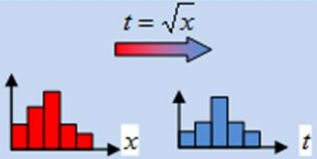
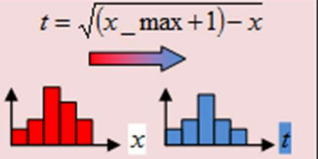
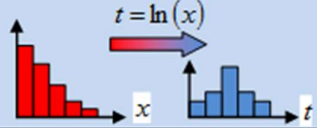
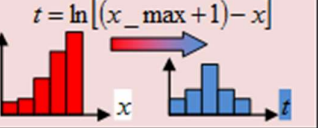
c. reprezentarea grafica a distributiei frecventelor absolute experimentale ale continuturilor de NH₄ – **HISTOGRAMA ABSOLUTA EXPERIMENTALA**

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



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

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

TRANSFORMARI PENTRU NORMALIZAREA DISTRIBUTIEI VALORILOR		
INTENSITATEA ASIMETRIEI	ASIMETRIE	
	POZITIVA	NEGATIVA
ZERO		
MODERATA	$t = \sqrt{x}$ 	$t = \sqrt{(x_{\text{max}} + 1) - x}$ 
SEVERA	$t = \ln(x)$ 	$t = \ln[(x_{\text{max}} + 1) - x]$ 
EXTREMA	$t = \frac{1}{x}$ 	$t = \frac{1}{(x_{\text{max}} + 1) - x}$ 

- **cantitativ**, utilizand coeficientul de asimetrie (SKEWNESS; =SKEW(v_i)):
 - $\beta_3 = 0$ – simetrie perfectă;
 - $\beta_3 > 0$ – asimetrie pozitivă sau “de stanga”;
 - $\beta_3 < 0$ – asimetrie negativă sau “de dreapta”.
- **Concluzie:** histograma absoluta experimentală a continuturilor de NH4 prezintă asimetrie pozitivă ($\beta_3 = +1,17$); în acest caz, **valoarea cea mai probabilă** a continuturilor de NH4 este **SUPRAESTIMATA** de valoarea medie a continuturilor de NH4.

e. **testarea concordantei** între repartitia **frecventelor absolute experimentale** $n_{i \text{ exp}}$ și **modelul repartitiei normale** – TESTUL χ^2

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

- **calculul frecventelor absolute teoretice** $n_{i \text{ teor}}$ (corespunzătoare modelului repartitiei 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));

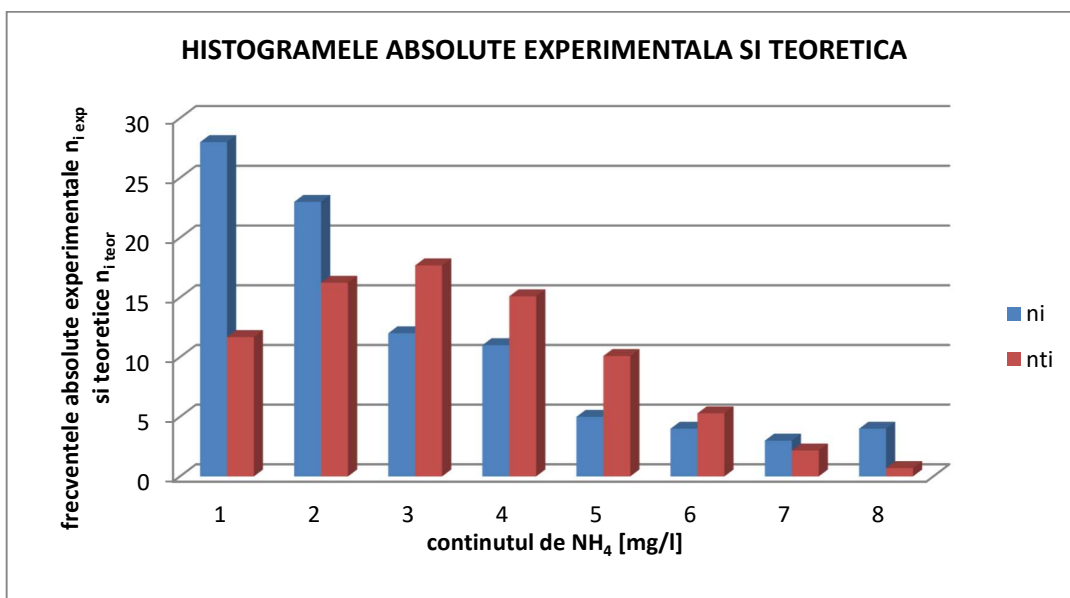
v_{ci} – valoarea centrala a fiecarui interval k de grupare;

$\bar{v}$ - valoarea medie a continuturilor de NH_4 (=AVERAGE(v_i)).

- **verificarea corectitudinii calculului frecventelor teoretice**

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

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



- **calculul statisticii** $\chi_{\text{experimental}}^2$

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

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

$$\chi_{\text{admisibil/teor}}^2 = f(\alpha, \nu)$$

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

m – numărul de parametri statistici utilizați în calculul frecvențelor absolute teoretice $n_{i\text{ teor}}$ (valoarea medie și abaterea standard; $m = 2$)

- **compararea statisticilor** $\chi^2_{\text{experimental}}$ **și** $\chi^2_{\text{admisibil/teoretic}}$
 - **dacă** $\chi^2_{\text{exp}} < \chi^2_{\text{adm/teor}}$, repartiția frecvențelor experimentale ale continuturilor de NH₄ **este NORMALA** pentru **riscul asumat** $\alpha = 10\%$. Valoarea cea mai probabilă a continuturilor de NH₄ este valoarea medie.
 - **dacă** $\chi^2_{\text{exp}} > \chi^2_{\text{adm/teor}}$, repartiția frecvențelor experimentale ale continuturilor de NH₄ **NU este NORMALA** pentru **riscul asumat** $\alpha = 10\%$. Valoarea cea mai probabilă a continuturilor de NH₄ NU este valoarea medie (20,99 mg/l) și este SUPRAESTIMATA de către aceasta. Se va încerca **NORMALIZAREA** distribuției frecvențelor continuturilor de NH₄.
 - **CONCLUZIE:** întrucât $\chi^2_{\text{exp}} (49,56) > \chi^2_{\text{adm/teor}} (9,24)$, **repartiția frecvențelor absolute experimentale ale continuturilor de NH₄ NU este NORMALA** pentru riscul asumat $\alpha = 10\%$.

f. evaluarea celei mai probabile valori în cazul unei selecții de date cu repartiție "ANORMALA"

- **normalizarea distribuției frecvențelor experimentale** utilizând transformări (prin intermediul unor funcții analitice) aplicate continuturilor de NH₄
 - **aplicarea transformărilor** 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 funcții analitice;
 - $\beta_{3T_1} = 5,48$
 - $\beta_{3T_2} = -1,55$
 - $\beta_{3T_3} = 0,36$
 - alegem **funcția radical** (T_3) pentru normalizarea frecvențelor continuturilor de NH₄

- **Concluzie:** aplicand metodologia A.V.G. si in urma aplicarii testului χ^2 , 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 NH4** 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(α, s, N)) depinde de:
 - riscul asumat (α);
 - 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 NH4** 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 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.