

---

## L7. SQL – Agregarea datelor.

### Extragerea informațiilor de sumarizare

Scopul agregării datelor este extragerea unor informații statistice, la nivel global, pentru a rezuma un set de date. În SQL funcțiile de agregare sunt funcții care au la intrare o colecție de valori și returnează o singură valoare relativă la datele primite. Cele mai multe funcții (exceptând **COUNT**, **MIN**, **MAX**) vor lucra folosind doar coloane sau expresii cu rezultat de tip numeric. Principalele caracteristici ale funcțiilor de agregare sunt:

- Operează pe o singură coloană sau o expresie pe un grup de coloane;
- Returnează o singură valoare pentru o grupare de rânduri;
- Pot fi folosit doar în **SELECT** și în clauza **HAVING**;
- în **SELECT** produc un rezultat anonim, deci se recomandă să fie redenumite.

#### 7.1. **COUNT**

Returnează numărul de valori pentru o coloană specificată sau pentru o expresie. Acesta va ignora valorile de **NULL** întâlnite (nu vor fi numărate). Utilizând specificatorul **DISTINCT**, funcția **COUNT** va număra câte valori distincte sunt primite, duplicatele vor fi considerate o singură dată. Dacă se utilizează **\***, atunci funcția **COUNT** va lua în considerare toate valorile din intrare.

Sintaxa generală este:

---

```
SELECT COUNT([DISTINCT] expresie) AS alias  
FROM Tabelă  
[WHERE condiție]
```



**ACTIVITATE 1:** Utilizând **SQL Workshop** -> **SQL Commands** rulați următoarea interogare SQL:

```
SELECT COUNT(*) AS nrStud,  
COUNT(bursa) AS nrBurse,  
COUNT(DISTINCT bursa) AS nrNiveluriBurse  
FROM Student;
```

## 7.2. **MIN și MAX**

Funcția **MIN** va primi o coloană sau o expresie și va returna valoarea minimă din mulțimea de valori primită. Funcția **MAX** va calcula valoarea maximă. Aceste funcții ignoră valorile nule, iar dacă toate valorile primite sunt nule ambele vor returna **NULL**. Valorile primite vor fi în general valori numerice, dar se acceptă și șiruri de caractere sau data calendaristică. În Oracle, compararea șirurilor de caractere ține cont, în mod implicit, dacă literele sunt majuscule sau minuscule.



**ACTIVITATE 2:** Utilizând **SQL Workshop** -> **SQL Commands** rulați următoarele interogări SQL:

(Nivelurile minime si maxime ale burselor)

```
SELECT MIN(bursa) AS bmin, MAX(bursa) AS bmax  
FROM Student  
WHERE bursa > 0;
```

---

### 7.3. **SUM și AVG**

Aceste funcții vor primi o coloană sau o expresie de tip numeric și vor returna suma, respectiv media aritmetică a valorilor primite. Aceste funcții ignoră valorile nule, iar dacă toate valorile primite sunt nule vor returna **NULL**.

Sintaxa generală este:

```
SELECT SUM(expresie1) AS aliasSUM,  
          AVG(expresie2) AS aliasAVG  
FROM Tabelă  
[WHERE condiție]
```

### 7.4. **VARIANCE și STDDEV**

Funcția **VARIANCE** returnează varianța unui set de valori pentru o coloană sau o expresie numerică. Varianța este media aritmetică a pătratului diferențelor valorilor față de mediană. Cu alte cuvinte măsoară cât de împrăștiate sunt valorile din set raportate la mediană. Funcția **STDDEV** extinde funcția anterioară, permițând obținerea unei informații mai specifice. Din punct de vedere matematic, aceasta calculează radicalul varianței. Rezultatul poate fi folosit pentru a defini intervalul de normalitate în jurul medianei unei mulțimi de valori. Aceste funcții ignoră valorile nule, iar dacă toate valorile primite sunt nule ambele vor returna **NULL**. Ambele valori vor returna zero dacă au la intrare o singură valoare nenulă.

Sintaxa generală este:

```
SELECT VARIANCE(expresie1) AS aliasVar,  
          STDDEV(expresie2) AS aliasStdDev  
FROM Tabelă  
[WHERE condiție]
```



**ACTIVITATE 3:** Utilizând **SQL Workshop** -> **SQL Commands** rulați următoarele interogări SQL:

---

(Varinața și intervalul de normalitate [m1..m2] a mediilor)

```
SELECT trunc(variance(media),2) AS varianta,  
trunc(avg(media),2) - trunc(stddev(media),2) AS m1,  
trunc(avg(media),2) + trunc(stddev(media),2) AS m2  
FROM Student;
```

## 7.5. Gruparea datelor

Gruparea datelor se folosește atunci când este nevoie să se obțină sumarizare/statistici pentru anumite grupuri de valori. Clauza **GROUP BY** este folosită în **SELECT** pentru a grupa rezultatele și produce câte un rezultat cu valori statistice pentru fiecare grup. În acest caz, lista de proiecție trebuie să conțină doar funcții de agregare sau câmpuri unice pentru fiecare grup. Clauza **GROUP BY** poate conține mai multe criterii de grupare, dar acestea nu vor fi ierarhizate. După grupare, rezultatele pot fi ordonate (ierarhizate) în continuare folosind **ORDER BY**. Dacă este cazul, se va crea un grup special pentru valorile nule. Este imperativ ca toate câmpurile libere (care nu apar în funcții de agregare) din lista de proiecție să fie incluse în lista de câmpuri din **GROUP BY**. Clauza **HAVING** poate fi folosită pentru a restricționa grupurile care vor apărea în rezultatul final.



**ACTIVITATE 4:** Utilizând **SQL Workshop** -> **SQL Commands** rulați următoarele interogări SQL:

```
SELECT an AS anStudiu, COUNT(*) AS nrStudAn  
FROM Student  
GROUP BY an  
HAVING COUNT(*) > 3;
```

---

## 7.6. **Exerciții**

**Obs:** în cazul în care nu se precizează câmpurile dorite în rezultat se vor afișa cele mai relevante câmpuri care permit identificarea înregistrării (ex. la Student: sid, nume și an).

**L7.Ex1.** Să se calculeze numărul de studenți care nu au bursă și au media peste 7.50.

**L7.Ex2.** Să se calculeze numărul de studenți care au ore cu profesorul 'Luca Caragiale'.

**L7.Ex3.** Să se calculeze numărul total de studenți care au ore la etajul 1 (încărcarea sălilor de la etajul 1), indiferent de facultate și de zi.

**L7.Ex4.** Să se afișeze data nașterii celui mai tânăr și celui mai bătrân student din anul 2.

**L7.Ex5.** Să se calculeze suma și media burselor pe fiecare an de studiu la fiecare facultate. Se vor afișa numele facultății, anul de studiu, suma și media burselor.

**L7.Ex6.** Să se afișeze bursa minimă ( $>0$ ), medie și maximă a tuturor studenților care au ore în ziua de 'Luni', indiferent de curs.

**L7.Ex7.** Folosind subinterogări și agregări să se afișeze toți studenții cu bursa peste bursa medie din universitate.

**L7.Ex8.** Să se calculeze intervalul de normalitate a mediilor pentru fiecare facultate în parte. Se va afișa numele fiecărei facultăți precum și pragurile de jos și de sus a intervalului de normalitate pentru mediile studenților din acea facultate.

**L7.Ex9.** Folosind subinterogări și agregări să se afișeze toți studenții care au medii în afara intervalului de normalitate a mediilor pe universitate.