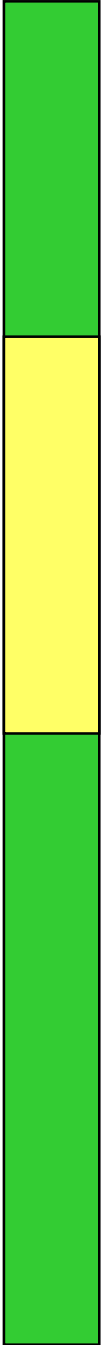


Tipuri de variabile, indicatori de  
tendință centrală și dispersie



Nu tot ceea ce poate fi măsurat are  
valoare și nu tot ce are valoare poate  
fi măsurat.

Bradley & Field 1995, EBM

# Tipuri de date

## Calitative (categorice)

- **Dichotomice** (masc./fem., succes/eșec),
- **Nominale** (grupă sanguină, stare civilă),
- **Ordinale** (anxietate, stadii ale cancerului, satisfacție).

## Cantitative

- **Discrete** (numărul de vizite la un medic de familie într-un an, numărul de episoade de boală în ultimii 5 ani),
- **Continue** (vârstă, tensiune arterială, greutate la naștere).

# Calitative

**Dichotomică (Binară):** Această variabilă are doar 2 categorii posibile (care se exclud una pe alta)

**Ex.** Rezultat: Succes/Eșec,  
Sex: Masculin/Feminin.

**Nominală:** Această variabilă are mai mult de 2 categorii, care se exclud una pe alta și care **nu** sunt ordonate

**Ex.** Grupele sanguine: A/B/AB/O,  
Stare civilă: Căsătorit/Divorțat/Necăsătorit/Văduv.

**Ordinală:** Această variabilă are mai mult de 2 categorii, care se exclud una pe alta și care sunt ordonate

**Ex.** Stadiul bolii: Ușoară/Moderată/Severă,  
Satisfacție: Foarte satisfăcut/Satisfăcut/Nesatisfăcut

# Cantitative

**Discretă:** Această variabilă reprezintă adesea “numărări” (valori întregi)

**Ex.** Numărul de vizite la medicul de familie într-un an,

Numărul de copii,

Numărul de internări la spital în ultimii 5 ani.

**Continuă:** Această variabilă ia orice valoare între anumite limite

**Ex.** Înălțimea în cm,

Greutatea în kg,

Distanța de acasă la locul de muncă în km.

# Ordinale vs. Discrete

## Ordinale

*Stadii în cancerul mamar*      I    II    III    IV

⇒ Spațierea nu este egală

## Discrete

*Nr. de copii*                    0    1    2    3    4    5+

⇒ Diferențele sunt egale

# Discrete vs. Continue

## Discrete

**Nr. de copii**                      0    1    2    3    4    5+

⇒ Numărare, valori numerice întregi

## Continue

**Greutate la naștere (kg)**            de la cca 1,5 la cca 4

⇒ Gamă de valori continue neîntrerupte

**Greutate la naștere**

<1.5    (1.5, 1.99)    (2, 2.49)    (2.5,2.99)    (3, 3.49)    (3.5, 3.99)    ≥4

# Importanța tipului de date

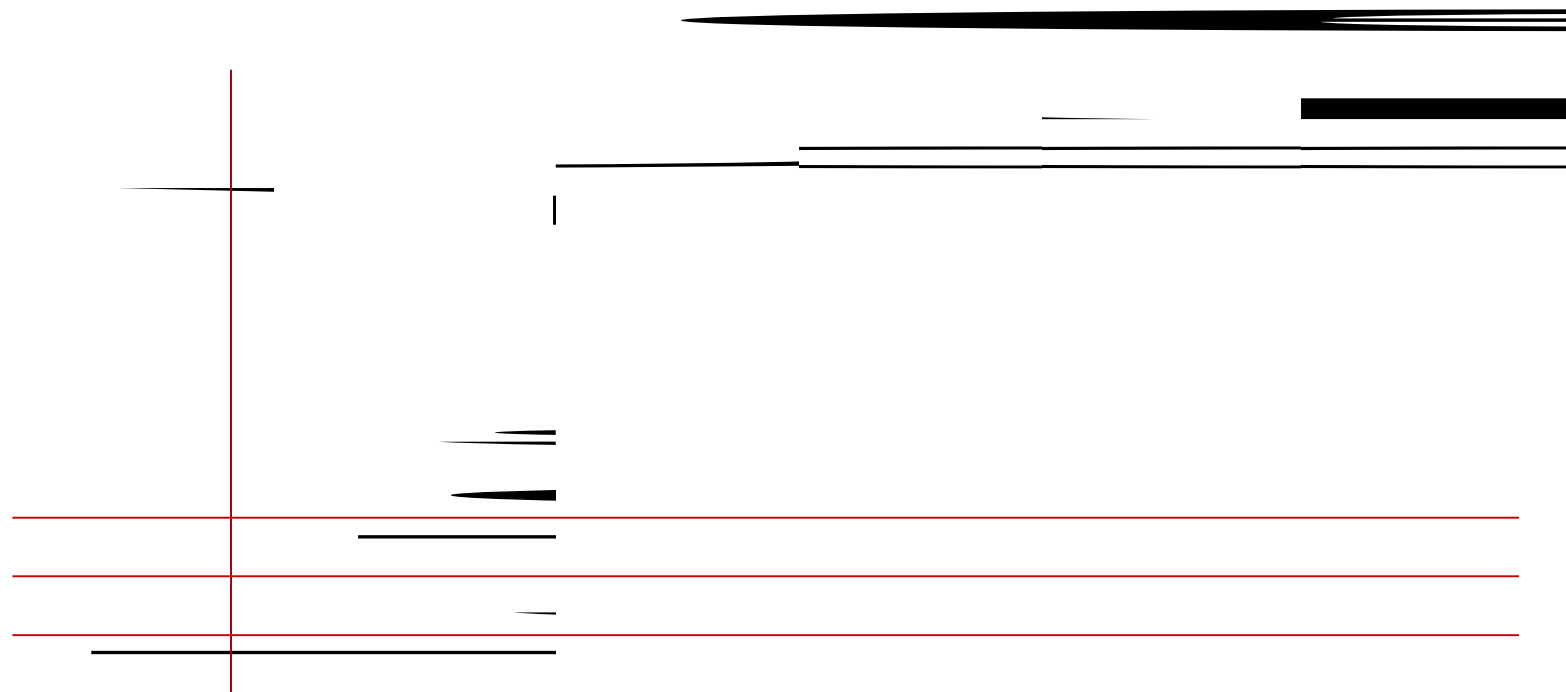
- ✚ Tipul de date are o importanță critică în determinarea metodelor de analiză potrivite și valide.
- ✚ Distincția majoră este între variabilele continue și cele categorice.

Verificarea sănătății	Cum se măsoară	Tipul de variabilă
Tensiunea arterială	mmHg	continuă
Colesterol	mmol/l	continuă
Diabet	Da/Nu	dichotomică
IMC (indice masă corp)	kg/m <sup>2</sup>	continuă

Caracteristici	Cum se măsoară	Tipul de variabilă
Vârstă	ani	continuă
Sex	fem./masc.	dichotomică
Slujbă	angajat/ pens./casnică (...)	nominală
Statutul de fumător	fumător, nefumător, ex-fumător	nominală
Exerciții fizice: cât de des?	ore pe săptămână	discretă
Consum alcool: cât de des?	unități pe săpt.	discretă
Consum alcool: de ce?	(numiți motive)	nominală
Nivelul sănătății	De la excelent (1) la îngrozitor (5)	ordinală

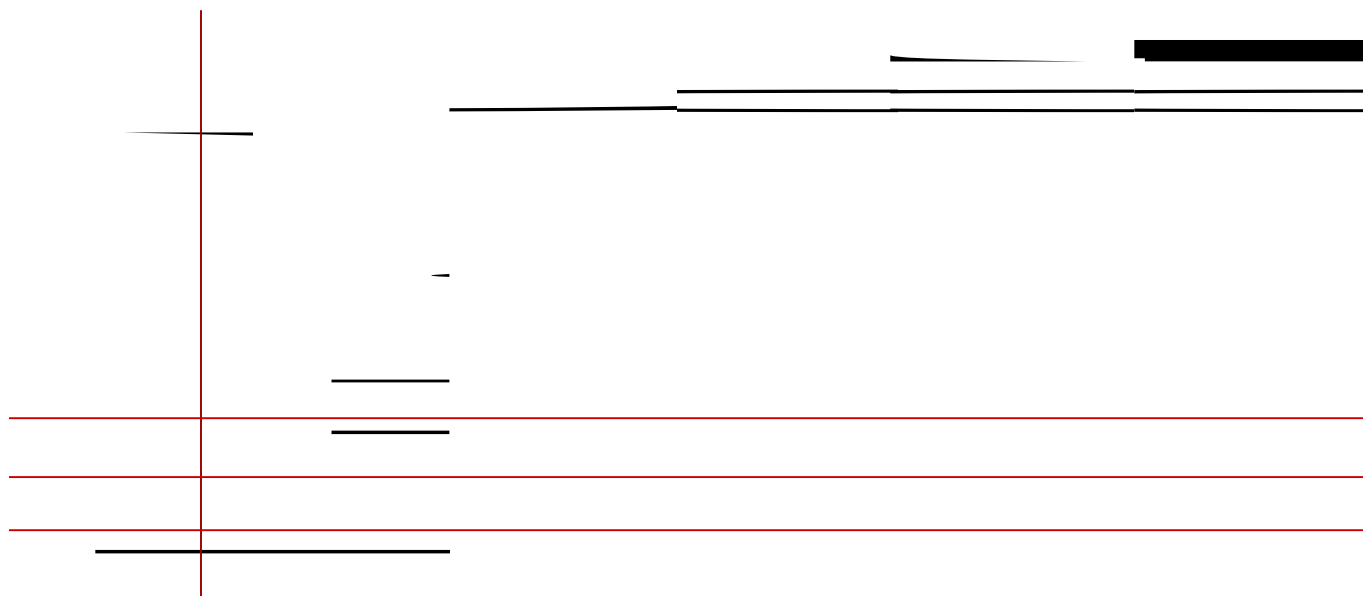
# Prezentarea datelor Nominale

⇒ Tabel de frecvențe



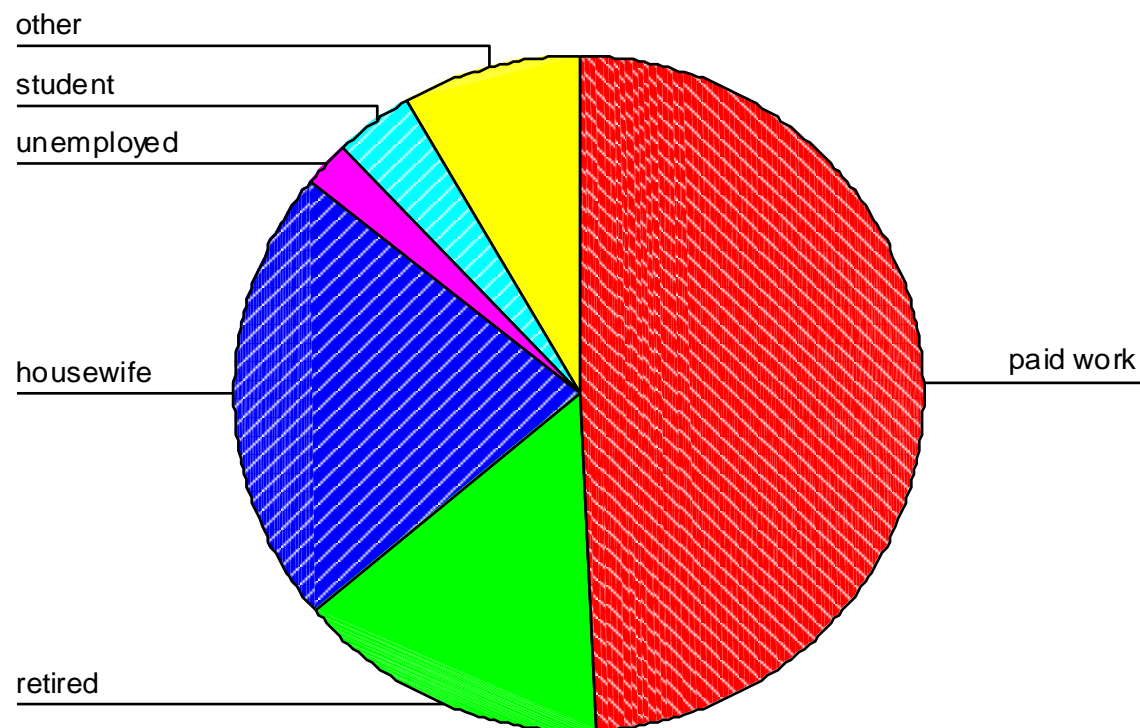
# Prezentarea datelor Ordinale

⇒ Tabel de frecvențe



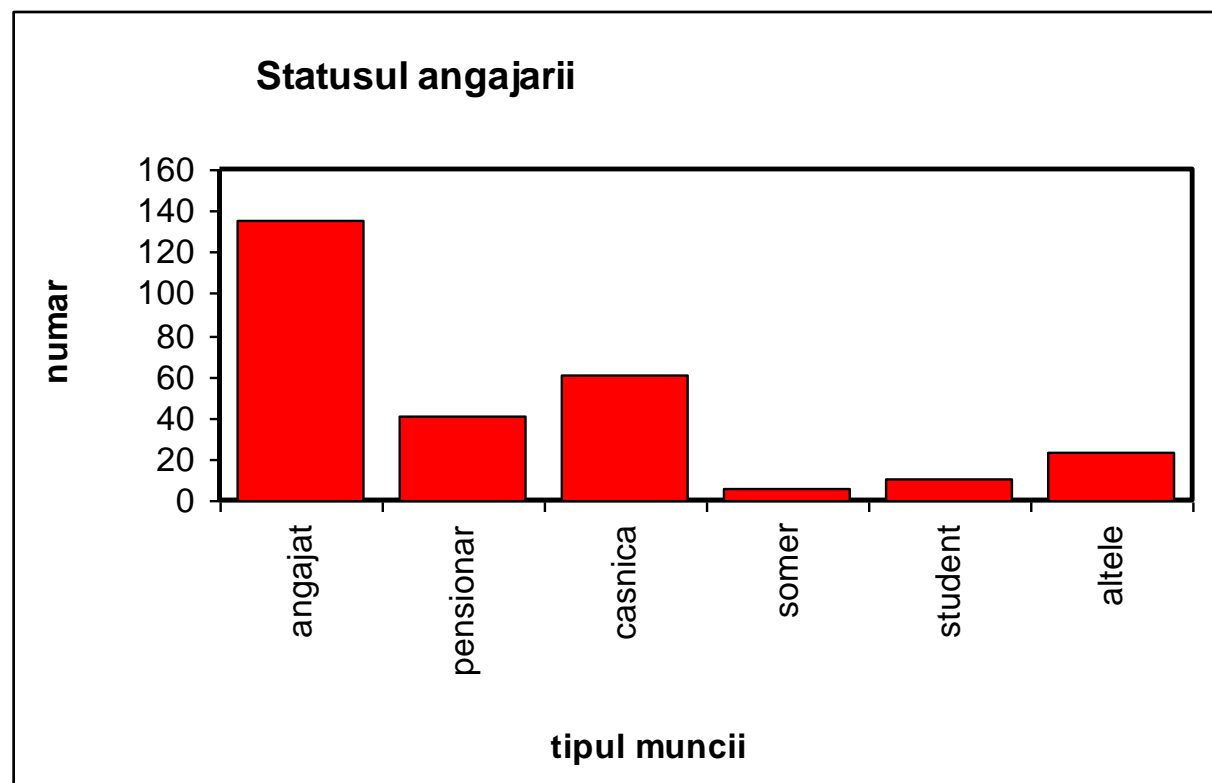
# Prezentare grafică a datelor calitative

Plăcintă ⇐



# Prezentare grafică a datelor calitative

Coloane ↵



# Două variabile categorice prezentate împreună

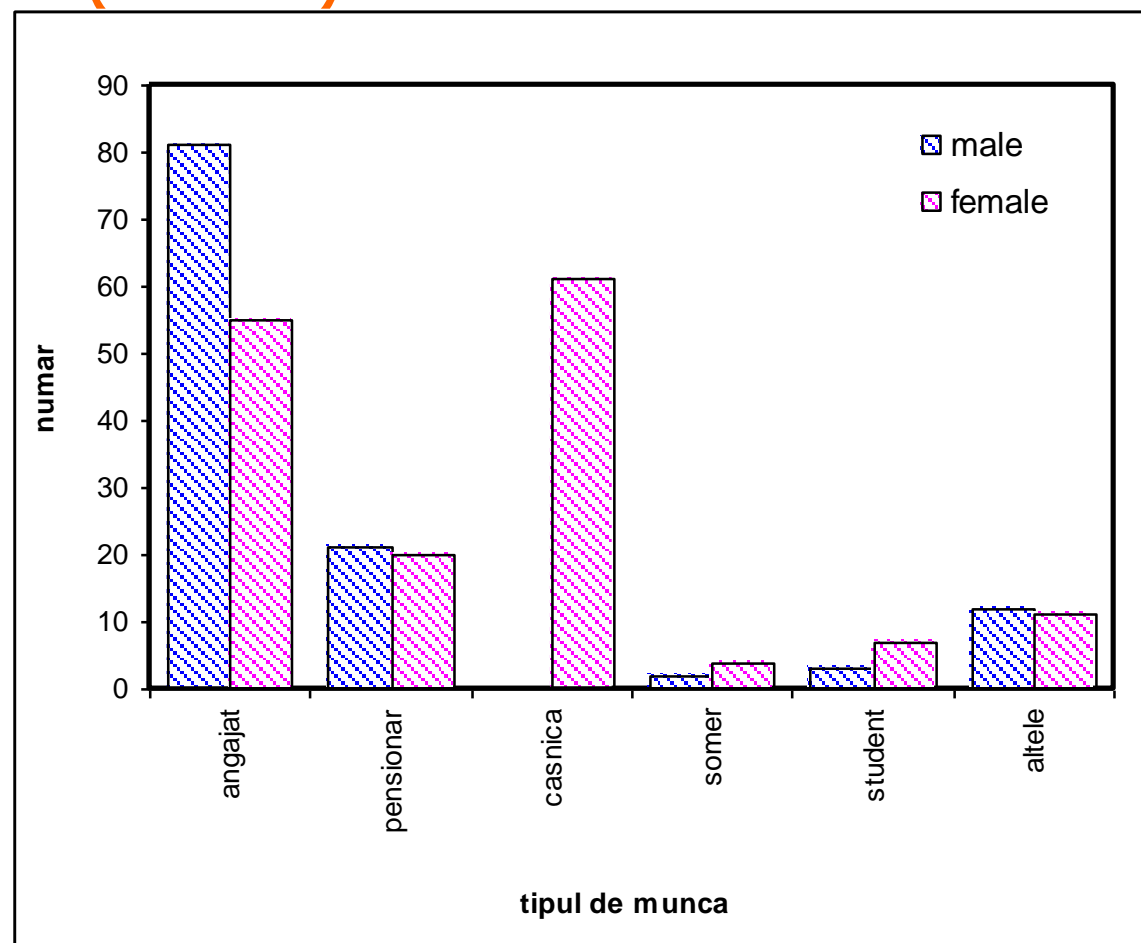
Crosstabulation: ocupație & sex ⇄

		gender of the subject		Total
		male	female	
type of work	paid work	81 68.1%	55 34.8%	136 49.1%
	retired	21 17.6%	20 12.7%	41 14.8%
	housewife		61 38.6%	61 22.0%
	unemployed	2 1.7%	4 2.5%	6 2.2%
	student	3 2.5%	7 4.4%	10 3.6%
	other	12 10.1%	11 7.0%	23 8.3%
Total		119 100.0%	158 100.0%	277 100.0%

# Reprezentare grafică 1

## Ocupație & Sex

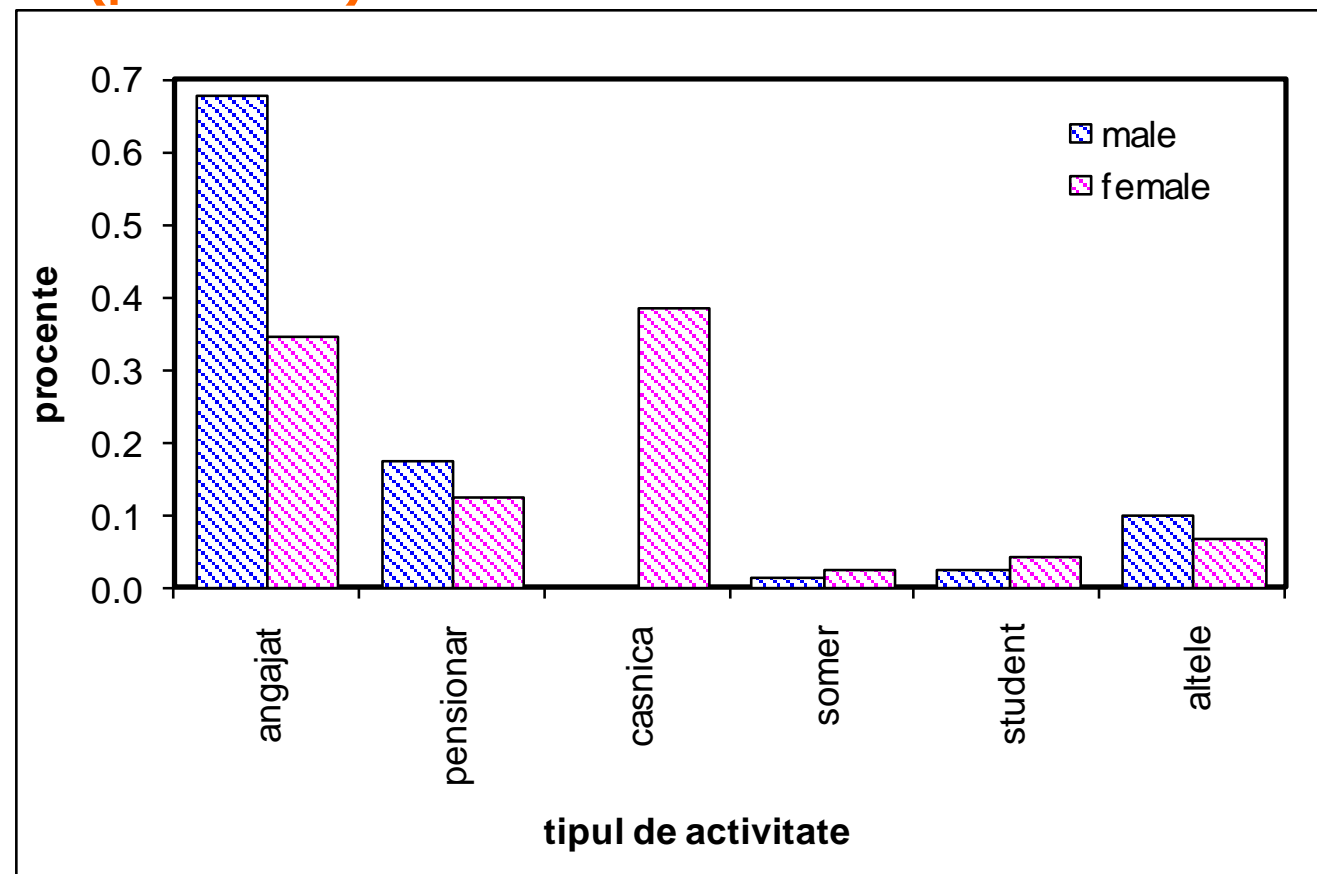
⇒ Coloane (numeric)



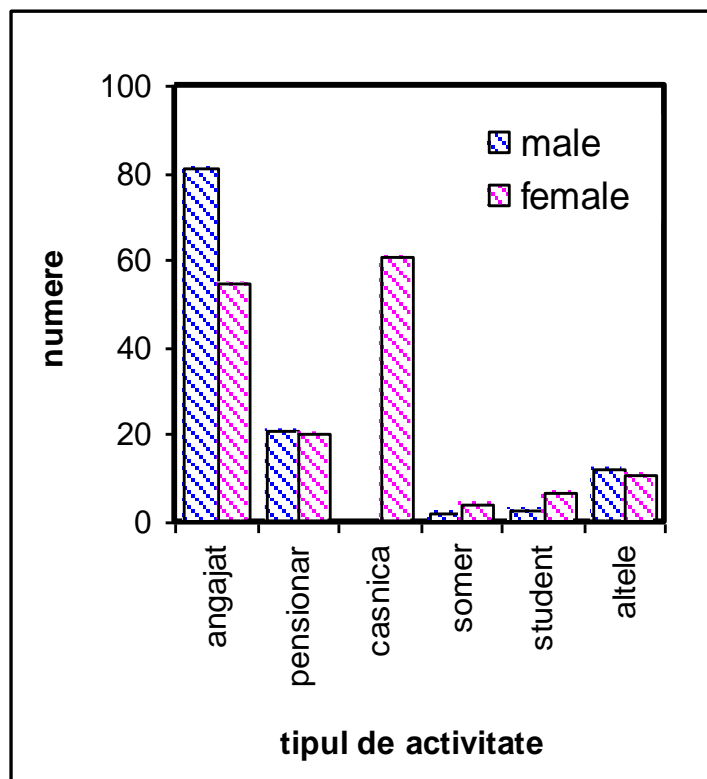
# Reprezentare grafică 2

## Ocupație & Sex

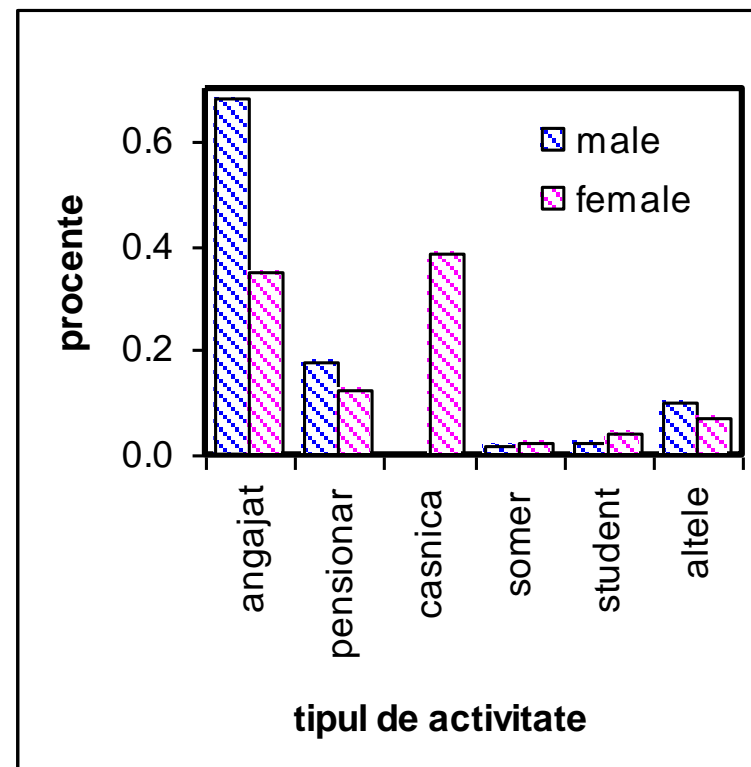
⇒ Coloane (procente)



# Frecvențe sau Procente?



bărbați n=119



femei n=158

# Prezentarea datelor calitative

Variabila	Screening pas cu pas (n= 1513) n (%)	Screening cuplu (n= 330) n(%)
<u>Starea civila</u>		
Casatorit	1 190 (79)	269 (81)
Concubinaj	224 (15)	45 (14)
Necasatorit	99 (6)	16 (5)
Partener prezent la vizita antenatala?	840 (51)	188 (52)

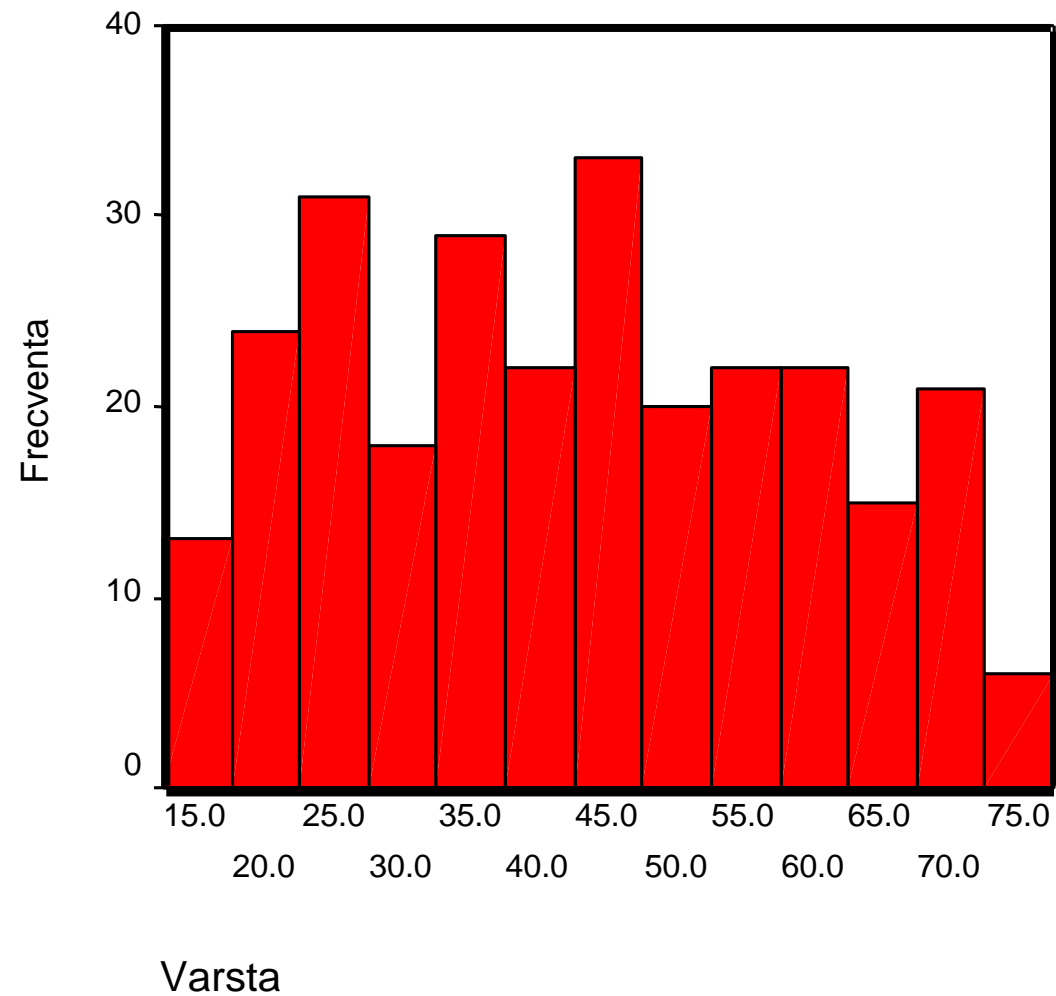
# Date Continue

✚ Pentru date continue este important să verificăm forma distribuției de frecvențe. Acest lucru se face cu ajutorul **histogramei**.

✚ **Forma distribuției** este importantă pentru determinarea celei mai potrivite măsurători care să indice locația și răspândirea datelor.

# Histograma

Distribuția de frecvențe ↵

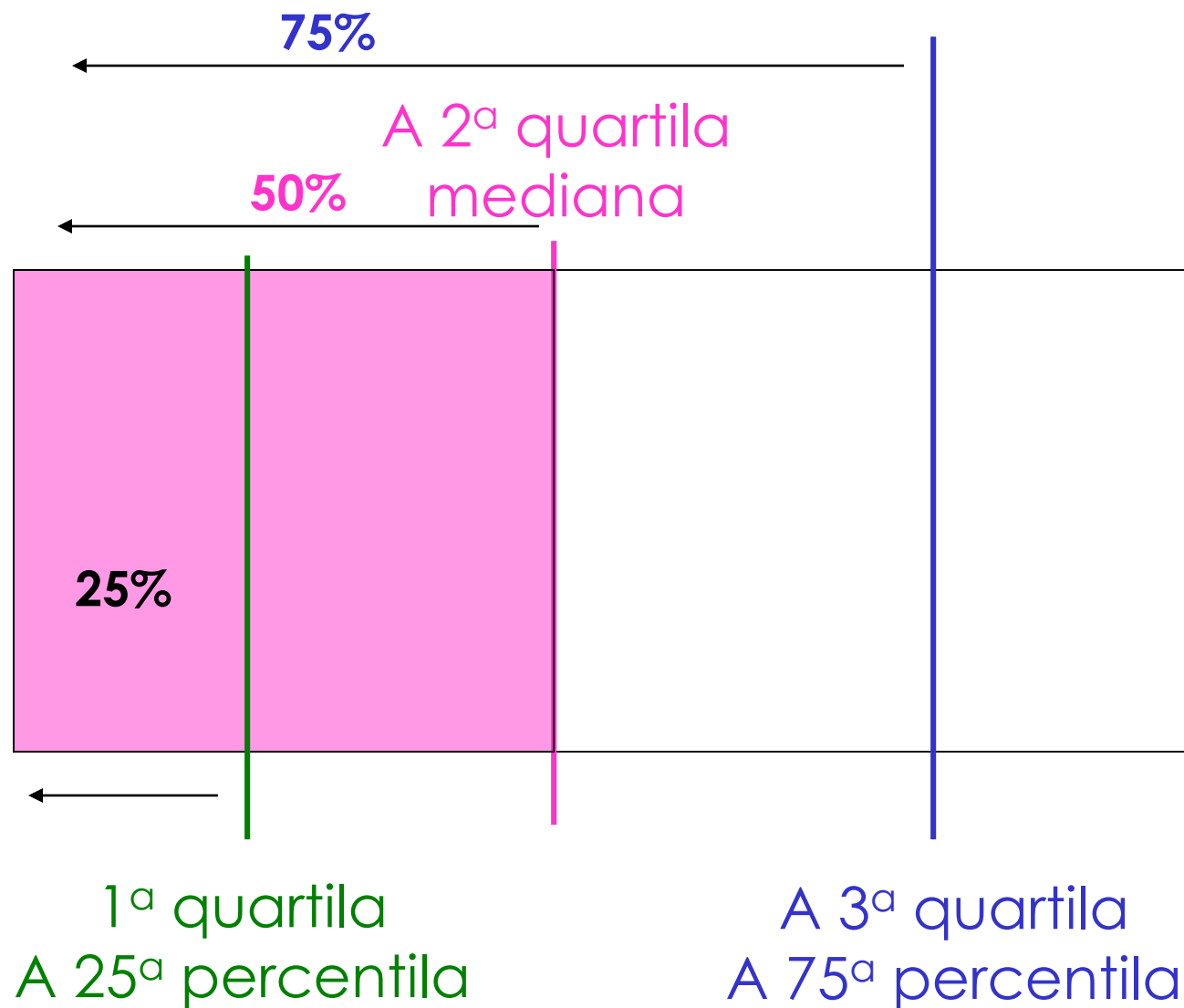


# Măsurători ale variației

Pentru a evalua variabilitatea observațiilor ⇐

- ✚ **Rangul**: valoarea maximă minus valoarea minimă,
- ✚ **Varianța eșantionului**: suma pătratelor distanțelor fiecărei observații față de medie împărțită la  $n-1$ ,
- ✚ **Deviația standard a eșantionului**: rădăcină pătrată (radical) din varianța eșantionului,
- ✚ **Coeficientul de variație**: deviația standard se exprimă ca procent al mediei,
- ✚ **Intervalul interquartilelor**: diferența între 25<sup>a</sup> percentilă și a 75<sup>a</sup> percentilă a setului de date.

# Quartile



## Exemplu: vârstele înregistrate la 282 subiecți

Media	43.1 ani	Mediana	43 ani
Minimum	15 ani	25 <sup>a</sup> percentila	25 ani
Maximum	75 ani	75 <sup>a</sup> percentila	56.3 ani
Rang	60 ani	IQR	31.3 ani
Varianța	280.5 ani <sup>2</sup>		
Deviația Standard	16.8 ani		
(DS)			

# Măsurători ale tendinței centrale

Indicatori ai centrului setului de date ⇐

⊕ **Media** - aritmetică

suma tuturor observațiilor și împărțirea lor la numărul de observații.

⊕ **Mediana** – valoarea centrală a distribuției

rangul observațiilor și mediana reprezintă observația sub care se găsesc 50% din toate valorile.

⊕ **Modul** – valoarea care se întâlnește cel mai des.

# Măsuri potrivite pentru tendențe centrale și dispersie

Tip de date	Central	Dispersie
Ordinale	Mediana	IQR
Variabile continue deviate	Mediana	IQR
Variabile continue simetrice	Media	Deviația standard

# Coeficientul de Variație

$$CV = DS/medie$$

- ✚ CV este o cantitate fără unitate de măsură care indică variabilitatea în jurul mediei în funcție de mărimea acesteia,
- ✚ Este mai bună decât DS pentru compararea gradului de variabilitate la diferiți subiecți, localizări, grupuri etc...
- CV este folosit adesea pentru a descrie eroarea de măsurare, adică variația între măsurătorile aceleiași cantități la același individ.

# Coeficientul de Variație

**Tabel 1.** Date de citometrie legate de martori și pacienți cu sferocitoză ereditară (SE)

	Martori (n=140)		SE severă (n=11)	
Parametru	Medie	DS	Medie	DS
RDW (g/l)	25.8	1.8	45.0	3.0
RDW: Reticulocyte Distribution Widths				

$$CV_{\text{controls}} = 1.8/25.8 = 0.0698 = 6.98\%$$

$$CV_{\text{severe}} = 3.0/45 = 0.0667 = 6.67\%$$

# Prezentarea datelor continue

EX. Vârsta înregistrată la 282 subiecți

## ⊕ **Medie, Deviație standard**

- Vârsta medie a respondenților a fost 43.1 ani (deviația standard (DS) = 16.8).

## ⊕ **Mediana, intervalul interquartilelor**

- Vârsta mediană a fost 43 ani (intervalul interquartilelor între 25 și 56.3).

# Prezentări grafice ale datelor continue

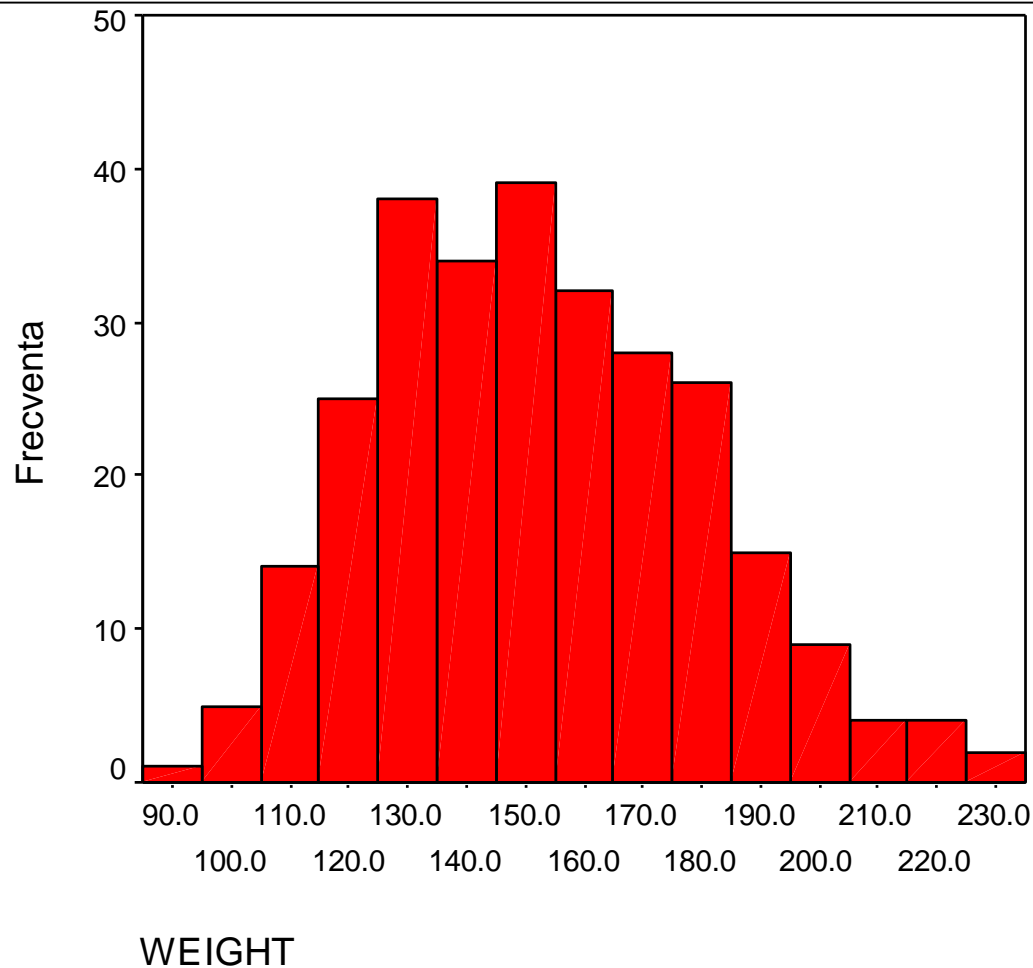
## ✚ Histograma

- indică forma distribuției de frecvențe.

## ✚ Box and Whisker Plot

- Indică forma și poate fi folositoare atunci când se compară două sau mai multe grupuri de subiecți.

# Histogramă pentru evaluarea normalității



⇒ Greutatea măsurată  
la 276 indivizi

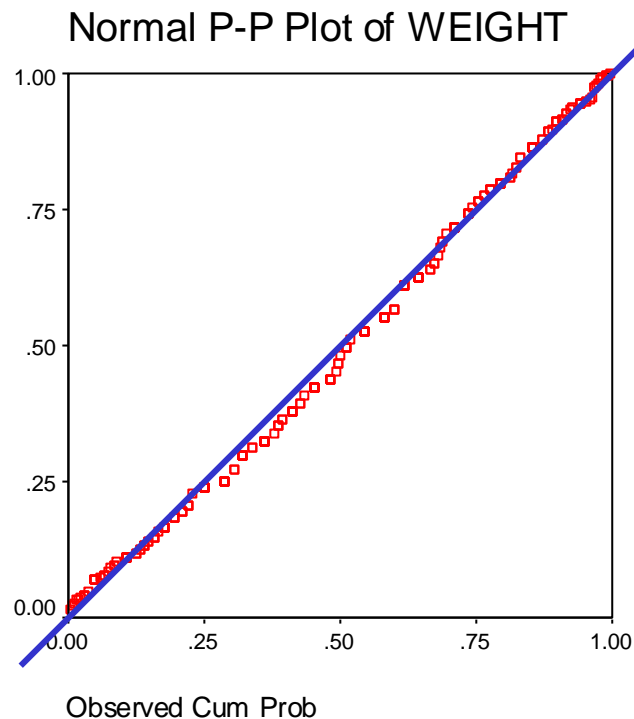
Std. Dev = 27.17

Mean = 152.4

N = 276.00

**Evaluarea normalității prin inspectarea vizuală a histogramei**

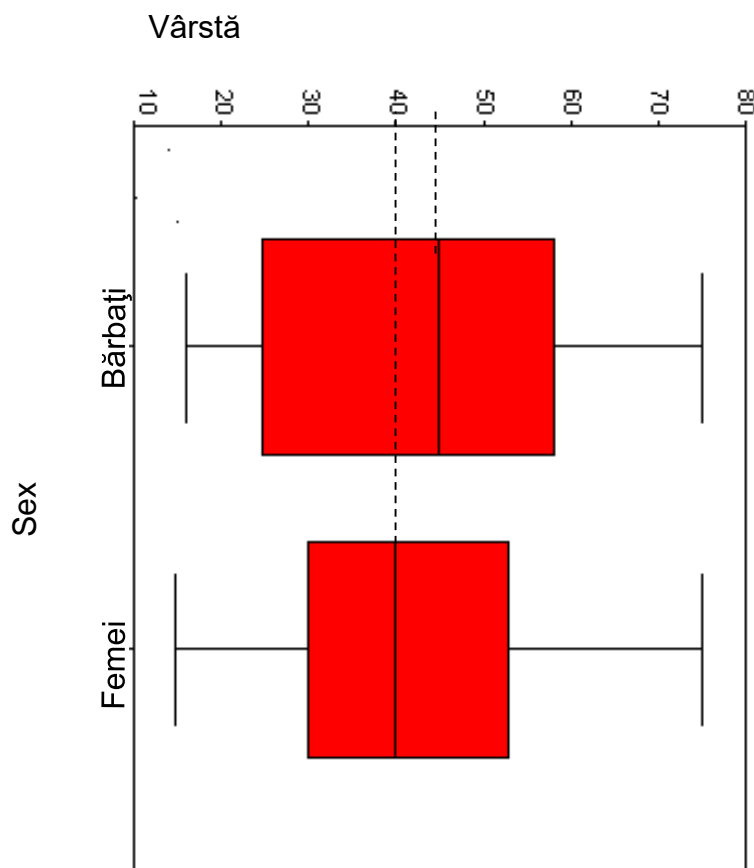
# Grafic de evaluarea normalității



- Acest grafic **vă ajută** să evaluați dacă se poate presupune că datele dvs. provin de la o distribuție normală
- Orice îndepărtare de la **linia dreaptă** este văzută ca o îndepărtare a datelor eșantionului de la normalitate

# Box and Whisker Plot

⇒ Vârsta a 276 indivizi (118 bărbați și 158 femei)



## Box and whisker plot:

- **Mediana** – linia orizontală din dreptunghi
- **25a și 75a percentilă** marginile dreptunghiului
- **Minimum și maximum** mustățile
- **Valorile extreme** în afara mustăților