## 1.1. Aprendre a obtenir mides de centralització i dispersió en una distribució estadística:

**Objectius:**

1. Comprendre la noció de distribució estadística i la seua no dependència de les propietats específics d'individus específics.
2. Aprendre a comparar diferents distribucions estadístiques pel valor al voltant del qual s'agrupen els seus valors.
3. Aprendre a comparar diferents distribucions estadístiques per la dispersió dels seus valors.
4. Entendre en quina mesura varien les mesures de centralització i dispersió d'una distribució estadística al sumar, restar, multiplicar o dividir els seus valors per una quantitat fixa.
5. Aprendre a calcular les mesures de centralització i dispersió de manera que se simplifiquen els càlculs i s'eviten els errors de cancelació.
6. Aprendre a normalitzar les distribucions estadístiques per tal de fer-les comparables més enllà de les seues mesures de centralització i dispersió.

**Activitat 1.2.** Una *variable aleatòria* (X) sobre un conjunt-població U és qualsevol variable que pot tenir distints valors (x) per als distints elements-individus de la població. La *distribución estadística* d'aquests valores no té en compte els individus concrets per als que aquesta variable té cada valor, sino quants la tenen, al que anomenem *frequència* f(x) d'aquest valor en la població. Anomenarem *paràmetre poblacional* a qualsevol quantitat que solament depenga de les frequències. Dues variables aleatòries seran estadísticament *equivalents* quan tinguen la mateixa distribució de freqüències.
Exercici 1.1: prendre una variable aleatòria sobre l'alumnat assistent a la classe, per exemple el fet de portar o no portar ulleres, i realitzar un experiment senzil per tal de comprovar que el número dels que porten ulleres és un paràmetre poblacional.

**Activitat 1.3.**
Exercici 1.2: representar gràficament en diagrames de barres la distribució estadística de les freqüències del número de calcer i de l'edat en l'alumnat assistent a classe.

**Activitat 1.4.** Com a mesures de *centralitat* (valor al voltant del qual s'agrupen els valors de la variable aleatòria) podem prendre:
La *moda*: aquell valor que tinga la màxima freqüència en la població.
La *mediana*: suposant que el conjunt de valors de la variable aleatòria esté ordenat, serà un valor que tinga tants individus amb un valor inferior com amb un valor superior.
La *mitjana* $\mu(X)$: suposant que els valors de la variable aleatòria siguen números reals, i que la *grandària* (número d'individus n(U)) de la població siga finit, ve donada per la suma dels valors $X_i$ per a tots els individus i de la població dividida per la seua grandària, $\mu(X) = \sum_i X_i / n(U)$ .
Teorema 1.1: $\sum_x f(x) = n(U)$ , $\mu(X) = \sum_x x \cdot f(x) / n(U)$ .
Problema 1.1: calcular les diferents mesures de centralització per a les distribucions

estadístiques de l'Activitat 1.3. Com podem utilitzar les freqüències per simplificar els càlculs?

**Activitat 1.5.** Per justificar que el càlcul de la mitjana és una operació lineal, demostrar els següent teoremes:
Teorema 1.2: si tenim dues variables aleatòries X, Y amb valors numèrics reals sobre la mateixa població U, $\mu(X+Y)=\mu(X)+\mu(Y)$ .
Teorema 1.3: si tenim una variable aleatòria sumable X y un número real constant c, $\mu(c\cdot X)=c\cdot\mu(X)$ .

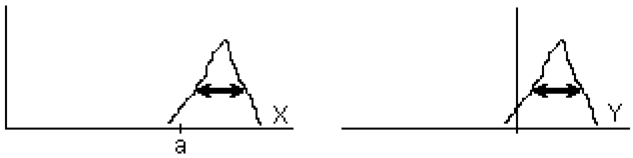
**Activitat 1.6.** A partir de la linealitat del càlcul de la mitjana expressada als dos teoremes anteriors, i tenint en compte que la mitjana d'una constant és igual a la mateixa constant, demostrar
Teorema 1.4: si tenim una variable aleatòria X amb valors numèrics reals i un número real $a\in R$, aleshores $\mu(X)=a+\mu(X-a)$ .
Teorema 1.5: si tenim una variable aleatòria X amb valors numèrics reals i dos números reals $a,c\in R$, i prenem $Y=(X-a)/c$, aleshores $\mu(X)=a+c\cdot\mu(Y)$ .
Els teoremes anteriors es poden utilitzar per a simplificar el càlcul de la mitjana. Aplicar-ho per a la resolució del
Problema 1.2: midar la longitud de la pròpia mà amb una precisió de 0'5 cm i calcular la mitjana del conjunt de l'alumnat assistent a classe.

**Activitat 1.7.** Com a mesures de *dispersió* (per expressar l'allunyament entre sí dels valors d'una variable aleatòria) podem prendre:
Els *quartils* primer i tercer: suposant que el conjunt de valors de la variable aleatòria esté ordenat, els quartils seran tres valors que dividisquen al conjunto de valors en quatre subconjunts de valors que corresponguen al mateix número d'individus; observem que el segon quartil coincidirà amb la mediana. Si tenim definida una distància en el conjunt de valors, podem mesurar la dispersió com la distància entre el primer i el tercer quartil.
La *amplitut*: suposant que els valors estiguen ordenats i tinguem definida una distància entre ells, serà la distància entre els valors mínim i màximo en la població.
La *desviació mitjana*: suposant que els valores de la variable aleatòria siguen números reals i que la grandària de la població siga finita, serà la mitjana del valor absolut de las diferèncias entre el seu valor per a cada individu i la mitjana d'aquests valors, $\mu(|X-\mu(X)|)$
La *variança* $\sigma^2(X)$: suposant que els valors de la variable aleatòria siguen números reals i que la grandària de la població siga finita, serà la mitjana del quadrat de les diferències entre el seu valor per a cada individu i la mitjana d'aquests valors, $\sigma^2(X)=\mu((X-\mu(X))^2)$.
La *desviació típica* $\sigma(X)$: es l'arrel quadrada de la variança.
Demostrar el
Teorema 1.6: $\sigma^2(X)=\mu(X^2)-\mu(X)^2$ (la variança és igual a la mitjana dels quadrats menys el quadrat de la mitjana).
Aquest teorema proporciona una forma més còmoda de calcular la variança.
Problema 1.3: calcular les diferents mesures de dispersió per a les distribucions estadístiques de l'Activitat 1.3.

Problema 1.4: calcular la variança d'aquest conjunt de valors: (1000000'1, 1000000'2, 1000000'2, 1000000'3).

**Activitat 1.8.** Com haurem vist a l'intentar resoldre el Problema 1.4, si la mitjana d'una distribució estadística és molt més gran que la seua amplitut, aleshores la mitjana del quadrat i el quadrat de la mitjana tindran moltes xifres significatives coincidents, que poden fins i tot superar la precisió dels nostres instruments de càlculs; en aquest cas, obtindriem erròniament zero com la seua diferència, produint-se així un "error de cancelació". Per tal de poder evitar-ho utilitzant les propietats de la variança, demostrar el

Teorema 1.7: si tenim una variable aleatòria X amb valors numèrics reals i un número real $a \in R$, i prenem $Y=X-a$, aleshores $\sigma^2(Y)=\sigma^2(X)$ , és a dir, la variança és invariant davant traslacions, com es pot entendre fàcilment observant la figura adjunta.




Per tant, podrem evitar l'error de cancelació restant a tots els valors una quantidad fixa pròxima al seu valor mínim. Aplicar-ho a la resolució del Problema 1.4.

**Activitat 1.9.** Demostrar el

Teorema 1.8: si tenim una variable aleatòria X amb valors numèrics reals i dos números reals $a,c \in R^+$, i prenem $Y=(X-a)/c$, aleshores $\sigma(X)=c\cdot\sigma(Y)$ .

Aplicar-ho per simplificar la resolució del

Problema 1.5: calcular la variança de la distribució estadística del Problema 1.2.

**Activitat 1.10.** Per comparar la forma de distribucions estadístiques amd diferents mitjanes i variances podem transformar-les en altres distribucions estadístiques amb mitjanes i variances coincidents. Anomenarem així *normalització* de una variable aleatòria X al resultat de restar-li la seua mitjana i dividir la diferència per la seua desviació típica, $N(X)=(X-\mu(X))/\sigma(X)$ . Demostrar el

Teorema 1.9: $\mu(N(X))=0$ i $\sigma(N(X))=1$.

Exercici 1.3: Representar gràficament en la mateixa figura la normalització de les distribucions estadístiques de l'Activitat 1.3.