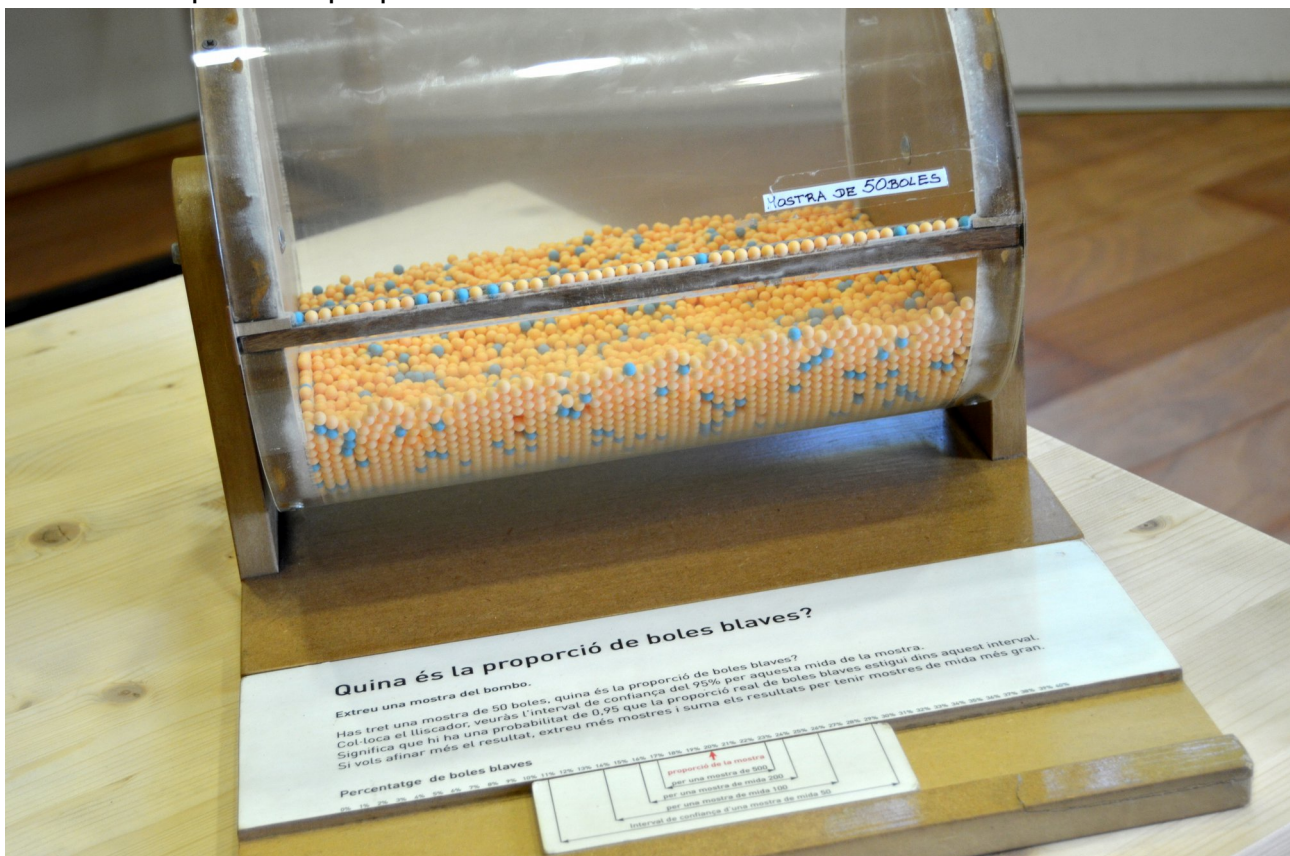


El bombo, les mostres i els intervals de confiança.

Aquest mòdul consisteix en un cilindre giratori transparent que conté aproximadament 2500 boletes de les quals un petit percentatge de color.

Quina és aquesta proporció de boles de color?

Un petit prestatge interior posat al llarg permet recollir, a cada gir, una mostra de 50 boletes. La mostra, doncs, es pot repetir fàcilment. Contant les boles de color de la mostra se sap la seva proporció de boles de color.



És clar que aquest percentatge, contat sobre la mostra, no té per què coincidir amb el percentatge global. Les matemàtiques, però, ens permeten determinar un interval de valors (interval de confiança) dins del qual quasi sempre (de mitjana 95 de cada 100 mostres) hi haurà el percentatge global.

Una peça lliscant que podem moure centrant-la en la proporció mostral observada, ens dona els extrems d'aquest interval de confiança.

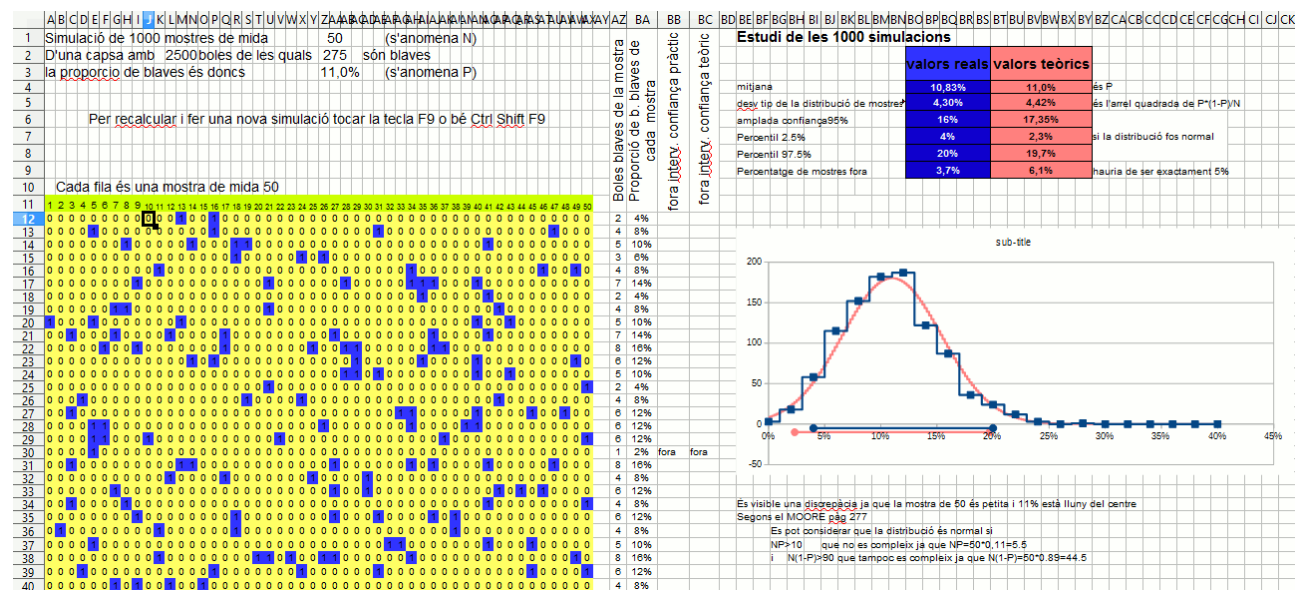
Podem repetir l'extracció de 50 boles i acumular-la amb les anteriors per obtenir mostres de mida 100, 200, 500. La peça lliscant mostra com, al augmentar la mida de la mostra, es redueix l'amplada de l'interval de confiança.

Amb aquest bombo al davant, no és gens complicat comprovar i entendre que la

repetició d'extraccions de mostres segueix unes pautes. Conèixer aquestes pautes ens permet quantificar la probabilitat que el resultat mostral estigui lluny del valor de la població.

Per aprofundir en aquest tema, podem seguir realitzant simulacions, però, utilitzant ara les eines tecnològiques. Una eina per fer-ho, que tothom té a l'abast, és el full de càlcul.

Al full que podeu descarregar a la web del **mmaca**¹ (en format Excel o en format LibreOffice) es simula l'extracció de 1000 mostres de mida 50 i es fa el tractament estadístic dels 1000 valors obtinguts (valors blaus).



Animació de 3 actualitzacions del full de càlcul

S'observa empíricament que la gràfica escalonada blava té sempre una mitjana al voltant de l'11% i una desviació típica del 4,4. Per aquests 1000 valors empírics es calcula la situació dels límits de l'interval que conté les 950 més properes a la mitjana.

Cada cop que s'actualitza el full de càlcul -amb les tecles F9 o bé amb la combinació de tecles Ctrl Shift F9 - les fórmules que contenen la instrucció ALEAT (RAND en anglès) generen uns nous valors i es visualitza una nova gràfica (de color blau) dels resultats de les 1000 noves extraccions.

No és ortodoxa considerar que en aquest cas la distribució de les mostres és una distribució normal. Així per exemple a Moore² (pàg. 277), es considera que cal que es compleixin les següents condicions:

$$NP > 10 \quad \text{i} \quad N(1-P) > 90$$

on N és la mida de la mostra i P la proporció de boles de color.

En el nostre cas no es compleix ja que $N=50$ i $P=0,11$,

1 http://www.mmaca.cat/images/mmaca/moduls/bombo_mostres [consulta set 2015]

2 MOORE, David S. **Estadística aplicada bàsica**. Antoni Bosch editor. Barcelona 1995

i per tant $NP=50*0,11=5.5$ i $N(1-P)=50*0.89=44.5$

Les necessitats de disseny de l'aparell han obligat a triar aquests dos valors. Calia que cada mostra fos simple de recomptar de forma visual. Una mostra de mida 50 és sempre una mostra molt petita, en situacions reals cal utilitzar mostres més grans i més si la proporció de boles de color no està al voltant del 50%, com és el nostre cas que està al 11% .

La distribució normal amb mitjana 0,11 i desviació típica 0,044 és la representada de color rosa. També es mostra l'interval de confiança calculat a partir de la normal.

Es visualitzen i comproven doncs, les discrepàncies entre els valors calculats segons cada simulació empírica de 1000 mostres (de color blau) i la distribució normal (de color rosa).

A la peça lliscant del mòdul, les amplades dels intervals de confiança estan calculades a partir de la normal. Tot i ser, doncs, estrictament incorrecte, ho hem fet així per simplicitat. Creiem que en aquest cas l'important és explicar el concepte d'interval de confiança.