

ÉCHANTILLONNAGE

QU'EST-CE QUE C'EST ?

BONJOUR,
JE SUIS ÉCHANTILLONNAGE

JE SERS À SÉLECTIONNER UN ÉCHANTILLON
REPRÉSENTATIF D'UNE POPULATION POUR
EN TIRER DES CONCLUSIONS STATISTIQUES

JE SUIS UTILE DANS DE NOMBREUX DOMAINES



STATISTIQUES
& SONDAGES



SCIENCES DES DONNÉES
& MACHINE LEARNING



TRAITEMENT
DU SIGNAL

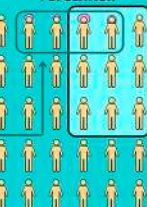
COMMENT ÇA MARCHE ?

DANS UNE POPULATION, JE SÉLECTIONNE
UN ÉCHANTILLON POUR EN TIRER
DES CONCLUSIONS STATISTIQUES

PAR EXEMPLE, SI JE M'INTÉRESSE
AUX CHEVEUX ROSES

P = PROPORTION DE PERSONNES
AUX CHEVEUX ROSES
DANS TOUTE LA POPULATION
ICI, $P = \frac{1}{10}$

POPULATION



ÉCHANTILLON



N = TAILLE TOTALE

F = PROPORTION DE PERSONNES
AUX CHEVEUX ROSES
DANS L'ÉCHANTILLON. ICI, $F = \frac{2}{10}$

SI ON CONNAÎT LE NOMBRE DE PERSONNES
AUX CHEVEUX ROSES DANS LA POPULATION, ON PEUT VÉRIFIER
L'INTERVALLE DE FLUCTUATION SUR L'ÉCHANTILLON

INTERVALLE DE FLUCTUATION

LORSQUE P EST CONNU ET QUE L'ON VEUT AVOIR DES INFORMATIONS
SUR L'ÉCHANTILLON, DANS ~ 95% DES CAS, F EST DANS L'INTERVALLE:

$$I = \left[P - \frac{1}{\sqrt{n}} ; P + \frac{1}{\sqrt{n}} \right]$$

INTERVALLE
DE FLUCTUATION
À 95%

TAILLE DE
L'ÉCHANTILLON

PROPORTION DANS LA POPULATION (EX: CHEVEUX ROSES)

EX: L'ENTREPRISE MONBREVET.COM PROMET 80%
DE RÉUSSITE AU BREVET DES COLLÈGES.

POUR VÉRIFIER
CETTE INFORMATION, J'INTERROGE
UN ÉCHANTILLON DE COLLÈGES.
SI LA MESURE DE F N'EST PAS DANS
L'INTERVALLE DE FLUCTUATION,
CETTE ENTREPRISE MENT !

SI ON NE CONNAÎT PAS LE NOMBRE DE PERSONNES
AUX CHEVEUX ROSES DANS LA POPULATION, ON VA SE SERVIR
DE L'ÉCHANTILLON POUR L'ESTIMER

INTERVALLE DE CONFIANCE

SI P EST INCONNU, ON PEUT LE DÉTERMINER À PARTIR D'UN
ÉCHANTILLON. DANS ~ 95% DES CAS, P EST DANS L'INTERVALLE:

$$I = \left[F - \frac{1}{\sqrt{n}} ; F + \frac{1}{\sqrt{n}} \right]$$

INTERVALLE
DE CONFIANCE
À 95%

TAILLE DE
L'ÉCHANTILLON

PROPORTION DANS L'ÉCHANTILLON (EX: CHEVEUX ROSES)

EX: JE CHERCHE À DÉTERMINER LES REVENUS >À 2000€/MOIS
DANS UNE POPULATION. J'INTERROGE DONC UN ÉCHANTILLON
ET JE DÉTERMINE F .

ON POURRA AINSI ENCADRER
CETTE VALEUR AVEC
L'INTERVALLE DE CONFIANCE.

CLOTILDE, STATISTICIENNE

JE M'APPELLE CLOTILDE ET JE SUIS
ENSEIGNANTE-CHERCHEUSE
ET STATISTICIENNE

DANS MON MÉTIER,
JE COLLECTE ET J'ANALYSE DES DONNÉES
POUR MIEUX COMPRENDRE COMMENT
FONCTIONNE NOTRE SOCIÉTÉ

THIS IS A MAN'S WORLD !

L'INSEE ME DEMANDE
UNE ÉTUDE POUR COMPARER LES SALAIRES
DES HOMMES ET DES FEMMES

AU LIEU D'APPELER
23 MILLIONS DE GENS,
NOUS ALLONS CRÉER
DES ÉCHANTILLONS.

TU PEUX LES APPELER ?

FLEMME !

J'AI APPELÉ 100 HOMMES ET 100 FEMMES.
RÉSULTAT: 51 HOMMES ET 53 FEMMES
GAGNENT PLUS DE 2000€

C'EST GÉNIAL !
ON Y EST !
C'EST L'ÉGALITÉ !!

OH, LALA !
TU VAS TROP VITE !
VÉRIFIE TON INTERVALLE
DE CONFIANCE !

AH OUI,
ON NE PEUT RIEN
CONCLURE AVEC
CES CHIFFRES...

ÉLARGIS
TON ÉCHANTILLON !

CE N'EST PAS
MOI QUI LE DIS
CE SONT LES MATHS
QUI LE PROUVENT !

MA FORMATION

BAC GÉNÉRAL (SPÉ MATHS) + LICENCE MATHS/STAT + MASTER ÉCO-STATS
+ DOCTORAT + QUALIFICATION CNU + RECRUTEMENT MCF (MAÎTRESSE DE CONFÉRENCES)

$$I = \left[0,51 - \frac{1}{\sqrt{100}} ; 0,51 + \frac{1}{\sqrt{100}} \right] = [0,41; 0,61]$$

$$I = \left[0,53 - \frac{1}{\sqrt{100}} ; 0,53 + \frac{1}{\sqrt{100}} \right] = [0,43; 0,63]$$

BON, OK...
CETTE FOIS-CI J'AI INTERROGÉ
1000 HOMMES ET 1000 FEMMES.
VOICI LES RÉSULTATS:

INTERVALLE	% REVENUS > 2000€
[0,55; 0,61]	entre 55% et 61%
[0,49; 0,55]	entre 49% et 55%

TON ÉCHANTILLON EST BIEN PLUS CONCLUANT:
NOUS NE SOMMES PAS ENCORE PARVENUS
À L'ÉGALITÉ DE SALAIRE ENTRE HOMMES ET FEMMES...