

L2 – Sociologie - Statistiques

TD – Estimation et échantillonage

Exercice 1 (Estimation de proportion). On pratique un référendum sur la population \mathcal{P} des adultes français, qu'on estime constituée de $N = 45\,000\,000$ individus. La variable X “réponse au référendum”, est un variable qualitative binaire, qui a deux modalités : “oui” et “non”. On cherche à connaître la proportion p de réponses “oui” dans la population \mathcal{P} . Après un sondage par téléphone auprès de $n = 1000$ adultes, 540 affirment voter “oui”.

A partir de ces informations, estimer la proportion p avec une confiance $c = 0.95$.

Exercice 2 (Estimation et comparaison de proportion). Un référendum est prévu. Un chercheur se demande si la proportion d'électeurs qui envisagent de voter “oui” est plus importante dans la ville A que dans la ville B . On note p_A et p_B ces deux proportions. Une enquête révèle que :

- (a) Sur un échantillon de $n_A = 400$ habitants de la ville A , 220 ont l'intention de voter “oui”.
- (b) Sur un échantillon de $n_B = 625$ habitants de la ville B , 300 ont l'intention de voter “oui”.

On demande :

1. Faire une estimation pour chacune des villes des proportions p_A et p_B du nombre d'habitants qui ont l'intention de voter “oui”.
2. Peut-on en déduire, avec une confiance de 95%, que l'une des proportions est significativement plus grande que l'autre ?

Exercice 3 (Estimation, échantillonnage, et comparaison de proportions et de moyennes). En sociologie, la **désirabilité sociale** est le biais qui consiste à vouloir se présenter sous un jour favorable à ses interlocuteurs. Ce processus peut être inconscient, ou au contraire être le résultat d'une volonté délibérée de manipuler son image aux yeux des autres ou de ne pas être stigmatisé socialement, d'être conforme aux attentes sociales. Ce concept est particulièrement pris au sérieux par les sociologues qui font des enquêtes. En effet ce biais peut-être dommageable quant aux résultats. Par exemple, l'*échelle psychométrique* de Marlowe-Crowne offre une mesure de la désirabilité sociable.

On considère donc la variable *Désirabilité Sociale* (D.S) exprimée sur une échelle normalisée allant de 0 à 20. On notera par \mathcal{P} la population des enfants. On désigne par X la variable égale au score D.S. sur cette échelle. Une étude a montré que pour la population globale \mathcal{P} , on a une moyenne $\mu \sim 10.5$ et un écart type $\sigma \sim 4.5$, et que cette variable suit une loi normale : $X \hookrightarrow \mathcal{N}(10.5; 4.5)$.

1.
 - (a) On choisit un enfant au hasard. Calculer la probabilité pour que son score D.S. soit supérieur à 13.
 - (b) Trouver la valeur de a telle que $\mathbb{P}[X \geq a] \sim 0.95$
2. Pour deux échantillons de 300 enfants et 150 adultes soumis au stress, les scores de désirabilité sociale sont donnés par le tableau suivant :

Classes	[2; 4[[4; 6[[6; 8[[8; 10[[10; 12[[12; 14[[14; 16[[16; 18[[18; 20[
Effectifs (enfants)	4	11	26	49	67	60	39	25	19
Effectifs (adultes)	1	8	13	30	35	37	17	7	2

En utilisant des estimations par intervalles de confiance, avec $c = 0.95$, peut-on dire qu'en situation de stress, les enfants ont un score D.S. moyen supérieur à celui des adultes ?

Exercice 4 (Estimation de moyenne). On s'intéresse cette fois à la taille de la population \mathcal{P} des adultes français. On suppose la taille moyenne m_e d'un échantillon de 1000 personnes prises au hasard est de 1.74m avec un écart-type observé de $\sigma_e = 2.1$. Estimer la moyenne de la taille de la population adulte française avec une confiance $c = 0.95$.

Exercice 5 (Estimation et comparaison de proportions). Un quotidien publie la quote de popularité d'un chef de gouvernement. Avril : 36% d'opinions favorables ; Mai : 38% d'opinions favorables. Le journaliste conclut que la cote de popularité du président a augmenté de 2%.

1. Sachant que ces proportions ont été calculées sur un échantillon de 1000 personnes, donner un intervalle de confiance de ces deux valeurs au niveau de confiance 0.95%.
2. Peut-on véritablement conclure que la côte de popularité du chef de gouvernement a augmenté ?

Exercice 6 (Estimation de moyenne, données par classes, INSEE 2019). L'échantillon suivant donne l'âge à l'accouchement de femmes françaises en 1994 :

ÂGE	[20, 25[[25, 30[[30, 35[[35, 40[[40, 45[[45, 50[[50, 55[[55, 60[[60, 65[
Effectifs	556	905	722,	201,	25	1	0	0	0

1. Calculer la moyenne, l'écart-type et l'écart-type corrigé de cet échantillon
2. Faire une estimation ponctuelle de l'écart-type de l'âge lors de l'accouchement en 1994
3. Faire une estimation de la moyenne d'âge à l'accouchement en 1994, avec une confiance de 95%.

Exercice 7 (*Estimation de moyennes, INSEE 2018*). Un échantillon de 2410 femmes donne une moyenne d'âge à l'accouchement en 2018 de $m_e = 30.5$ ans avec un écart-type $s_e = 2.3$ ans.

1. Faire une estimation de l'écart-type avec une confiance de 95%.
2. Faire une estimation de la moyenne d'âge à l'accouchement en 2018 avec une confiance de 99%.
3. Déterminer la taille de l'échantillon nécessaire pour avoir une précision de 0.1 ans pour une confiance de 99%.

Exercice 8 (*Estimation et comparaison de moyennes, INSEE 2017*). En 2016, un échantillon de 29 fonctionnaires français gagne un salaire mensuel moyen net de $m_e = 2230$ eur avec un écart-type de 21.2eur. Fin 2017, un échantillon de fonctionnaires donnent les résultats suivants :

Salaire mensuel	[2200,2220]	[2220,2240]	[2240,2260]	[2260,2280]	[2280,2300]	[2300,2320]	[2320,2340]	[2340,2360]
Effectifs	8	16	36	43	51	28	16	2

1. Estimer l'écart-type et la moyenne des salaires nets des fonctionnaires français en 2016 avec une confiance de 95%.
2. Estimer la moyenne des salaires nets des fonctionnaires français fin 2017, avec une confiance de 95%.
3. Quelle taille d'échantillon faudrait-il pour estimer cette dernière moyenne avec une précision de 1.
4. Peut-on déduire des deux estimations précédentes que les salaires moyens nets des fonctionnaires français ont augmentés entre 2016 et fin 2017.

Exercice 9 (*Estimation de moyenne et d'écart-type, petit échantillon, INSEE septembre 2019*). Un échantillon de ménages français donne, sur un indicateur standardisé (une sorte de note), la confiance qu'il a sur la situation économique du pays :

Note de confiance	100	101	102	103	104	105	106	107	108	109	110	111	112	113
Effectifs	1	2	1	1	3	1	3	0	1	1	0	1	0	0

1. Calculer la moyenne, l'écart-type et l'écart-type corrigé de cet échantillon.
2. Estimer l'écart-type des notes avec une confiance de 95%.
3. Estimer la moyenne des notes avec une confiance de 99%
4. Peut-on estimer que la confiance des ménages est en moyenne supérieure à 100 ?
5. Quelle taille d'échantillon aurait été nécessaire pour une précision sur l'estimation égale à 1 de la moyenne des notes ?