Préparation à l'agrégation externe de Sciences Sociales

Statistique inférentielle - Estimateurs et Intervalles de confiance

2022-2023

Exercice 1 (2009)

1. Lors des élections européennes, une des listes présentée dans le Grand Ouest a obtenu 32% des voix. Lors du dépouillement dans un bureau de vote de Nantes, on compte 925 bulletins.
   (a) On appelle X la variable aléatoire qui compte le nombre de bulletins en faveur de cette liste. Donner la loi de probabilité suivie par X et ses paramètres.
   (b) On admet qu'elle peut être approchée par une loi normale. Quels sont les paramètres de cette loi ?
   (c) Quelle est la probabilité que cette liste ait obtenu entre 30% et 40% des voix dans ce bureau ?
2. La tête de liste envisage de se présenter aux élections législatives de 2012. Pour donner du poids à sa candidature, elle fait effectuer un sondage. Sur 200 personnes, 46 se disent prêtes à voter pour elle.
   (a) Quelle estimation de son score peut-on lui proposer ? Quel est l'estimateur associé ?
   (b) Déterminer un intervalle de confiance à 95% de l'estimation précédente.

Exercice 2 (2015)

On cherche à estimer la proportion p inconnue de ménages possédant un bien d'équipement donné, puis son évolution au cours du temps. A cet effet, on réalise deux enquêtes. Pour la première enquête, on choisit au hasard et de façon indépendante n_1 foyers dans la population totale dont l'effectif est beaucoup plus grand. On appelle S_1 la variable aléatoire correspondant au nombre de foyers possédant le bien d'équipement dans ce premier échantillon. Pour la deuxième enquête, on choisit au hasard et de façon indépendante n_2 foyers (n_2 pouvant être différent de n_1). On appelle S_2 la variable aléatoire correspondant au nombre de foyers possédant le bien d'équipement dans ce deuxième échantillon.

1. Quelles sont les lois de S_1 et de S_2 Donner leur espérance et leur variance.
2. On définit les variables aléatoires F_1 et F_2 par $F_1 = \frac{S_1}{n_1}$ et $F_2 = \frac{S_2}{n_2}$. Calculer l'espérance et la variance de F_1 et F_2.
3. Montrer que F_1 et F_2 sont des estimateurs sans bias de p.
4. On pose $G = \frac{F_1+F_2}{2}$. Calculer l'espérance de G. Que peut-on en déduire pour G ?
5. Calculer la variance de l'estimateur G.
6. On suppose $n_1 > n_2$. A quelle condition G est-il un meilleur estimateur de p que F_1 et F_2 ?
7. De manière générale, on s'intéresse aux estimateurs de p de la forme $uF_1 + vF_2$. Déterminer une condition sur les coefficients réels u et v pour que $uF_1 + vF_2$ soit un estimateur sans biais de p et $uF_1 + vF_2$ soit de variance minimum parmi les estimateurs de cette forme. Interpréter le résultat obtenu.

Exercice 3 (2013)

Certains sujets abordés dans les enquêtes d'opinion sont parfois assez intimes, et on court le risque que les personnes interrogées se refusent à répondre franchement à l'enquêteur, faussant ainsi le résultat.On peut alors avoir recours à une astuce consistant à inverser aléatoirement les réponses. Considérons une question confidentielle pour laquelle on veut estimer la probabilité p de réponses positives. Imaginons que l'enquêteur demande à chaque personne interrogée de lancer un dé.

— Si le dé tombe sur 6, la personne doit donner sa réponse sans mentir,

— sinon elle doit donner l'opinion contraire à la sienne.

Comme l'enquêteur ignore le résultat du dé, il ne pourra pas savoir si la réponse est franche ou non, et on peut espérer que la personne sondée acceptera de jouer le jeu. Dans cet exercice, on généralise la situation en tirant pour chaque personne une variable de Bernoulli de paramètre α.

— Si le résultat de cette variable est 1, la réponse est franche,

— sinon, elle est inversée.

(dans le cas du dé, $\alpha = 1/6$) Soit n le nombre de personnes interrogées.

1. Montrer que la probabilité qu'une personne interrogée réponde «oui» à l'issue de la procédure est $q = \alpha p + (1-\alpha)(1-p)$
2. Montrer que F_n, la fréquence observée par l'enquêteur, est un estimateur sans biais de q et de risque quadratique tendant vers 0 quand n tend vers $+\infty$.
3. Pour $\alpha \neq 1/2$, exprimer p en fonction de q.
4. En déduire que $T_n = \frac{F_n - 1 + \alpha}{2\alpha - 1}$ est un estimateur sans biais de p dont le risque quadratique tend vers 0 quand n tend vers $+\infty$.
5. Pour n fixé, quelle valeur attribuer à α pour que le risque quadratique soit minimum ? Est-ce acceptable ? Pour quelle valeur de ce risque est-il maximum ? Quel sera le risque quadratique avec le dé ?