
Statistique descriptive

Examen blanc de décembre 2011

Question 1 (... points)

Comme chaque année, le Conseil des Recteurs (CRef) s'intéresse au taux de réussite en bachelier dans les universités francophones. Les étudiants concernés étant 50.000, il est plus rapide d'effectuer un sondage afin de donner une estimation de ce taux de réussite. On choisit alors 200 étudiants. Parmi ceux-ci, 110 seulement ont réussi.

- a. Quel est le taux de sondage ?
- b. Quelle taille d'échantillon faudrait-il considérer pour avoir une marge d'erreur inférieure à 3% ?

Comme le taux de réussite varie entre les différentes universités, on choisit ensuite 200 étudiants répartis dans ces différentes universités : Liège, Bruxelles, Louvain-la-Neuve et Namur. Au total, il y a 15.000 étudiants de bachelier à Bruxelles, 12.000 à Louvain-la-Neuve, 9.250 à Namur et 13.750 à Liège. Dans chacune de ces universités, on choisit respectivement 65, 40, 37 et 58 étudiants pour réaliser le sondage. Les estimations obtenues pour le taux de réussite dans chaque université sont : 0.4 à Bruxelles, 0.2 à Louvain-la-Neuve, 0.3 à Namur et 0.28 à Liège.

- a. Dans un sondage stratifié, qu'est-ce qu'une variable auxiliaire ? Quel est l'espoir du sondeur à l'égard de cette variable ?
- b. Quel est le taux de sondage dans chaque strate ? Que mesurent-ils ?
- c. Estimez, sans biais, la proportion d'étudiants de bachelier ayant réussi leur année à Bruxelles et calculez un intervalle de confiance 95% pour cette proportion.
- d. Donnez un intervalle de confiance 95% pour cette proportion mais, cette fois, au niveau de l'ensemble de la population.
- e. Quelles sont les tailles d'échantillon préconisées avec une allocation proportionnelle (pour un échantillon global de 200 individus) ? Détaillez vos calculs.
- f. L'institut de sondage étant situé à Liège, on estime le coût par individu interrogé à 20€ pour Liège, 35€ pour Namur, 40€ pour Louvain-la-Neuve et à 55€ pour Bruxelles. Combien d'individus pourra-t-on interroger dans chaque strate (toujours en allocation proportionnelle) avec un budget total de 30.000€ ?

Question 2 (... points)

Dans la faculté de sciences appliquées, une enquête est réalisée auprès de 40 étudiants de première année pris au hasard (20 garçons et 20 filles) afin de connaître leurs loisirs et plus particulièrement le temps (en minutes) qu'ils passent quotidiennement sur Facebook.

Voici les données récoltées :

- Données brutes pour les garçons
180 120 180 260 240 120 180 120 240 170 150 120 180 165 150 200 150 180 150 165

- Distribution empirique chez les filles

Nombre de minutes	30	45	60	75	90	120	150	200	240	300
Fréquence	3	1	3	1	3	4	1	1	2	1

- Pour cet échantillon, déterminez les quartiles des temps consacrés à Facebook séparément pour chaque sexe. Détaillez et justifiez vos réponses **(3 pts)**
- Dans cet échantillon, construisez un graphique permettant de visualiser le lien entre le temps passé sur Facebook (par jour) et le sexe. **(1.5 pts)**
- Calculez, séparément pour les filles et les garçons, la moyenne et les variances des temps passés sur Facebook. Détaillez vos calculs. **(3 pts)**
- De quel type sont les variables considérées dans cette enquête ? Quelle est la variable explicative et quelle est la variable réponse ? Justifiez.

Question 3 (... points)

Dans une enquête commanditée par l'INAMI, on cherche entre autres à savoir s'il existe un lien entre le genre et l'état de santé autoévalué (« bon », « moyen » et « mauvais ») des individus. Dans l'échantillon considéré ($n = 1778$), il y a :

- 874 hommes,
- 328 personnes qui estiment leur état de santé comme étant « moyen » ;
- 699 hommes qui estiment leur état de santé comme étant « bon » ;
- 151 hommes qui estiment leur état de santé comme étant « moyen » ;
- 52 femmes qui estiment leur état de santé comme étant « mauvais ».

Représentez graphiquement le lien éventuel qu'entretiennent ces deux variables dans l'échantillon (détaillez les différentes étapes de la procédure et interprétez le graphique).

Question 4 (... points)

Supposons que, sur une période de 24 heures, la distribution des vitesses des véhicules mesurée par un radar automatique soit normale de moyenne 102 km/h et d'écart-type 4,9 km/h.

- Quelle est la proportion de véhicules ne respectant pas la limitation de vitesse à 120km/h ?
- Dans quel intervalle (centré sur la moyenne) observe-t-on 70 % des vitesses mesurées ? »