LICENCE Eco Gestion 2$^{\text{ième}}$ année - Statistiques II

Année Universitaire 2018 - 2019

Feuille d'exercices n° 3 : **Echantillonnage**

I - Une population comprend 5 nombres : 2, 3, 6, 8 et 11.

1. Calculer la moyenne et la variance de cette population.
2. On considère tous les échantillons possibles de taille 2 tirés **avec remise** de cette population.
   a. Combien sont-ils?
   b. Calculer la moyenne et la variance de la distribution d'échantillonnage de la moyenne de ces échantillons.
3. Mêmes questions qu'au **2**) dans le cas d'un tirage **sans remise**.
4. Retrouver les résultats à l'aide des formules des distributions d'échantillonnage des moyennes.

II - Une urne contient cinq boules, trois blanches et deux noires. On se propose de tirer tous les échantillons possibles de taille trois, d'abord avec remise puis sans remise.

1. Calculer dans les deux cas, le nombre d'échantillons.
2. Calculer dans les deux cas, la moyenne et la variance de la distribution d'échantillonnage des fréquences.
3. Retrouver ces résultats à l'aide des formules des distributions d'échantillonnage.

III - En 1970, les 1200 locataires d'une tour d'habitation ont des poids répartis de la façon suivante :

Poids en Kg	Proportion des locataires
25	0,2
50	0,3
75	0,4
100	0,1

Chaque ascenseur de l'immeuble à une charge limite de 1,4 tonne.

1. Si 21 locataires quelconques se pressent dans un ascenseur, quelle est la probabilité qu'il ne soit pas en surcharge ? Toute la démarche permettant de parvenir au résultat sera développée. (on donne $525 = 25 * 21$).
2. En 2010, cette dernière probabilité vaut $\frac{2}{5}$. De quel pourcentage la population de cette tour a-t'elle grossi entre 1970 et 2010. On supposera que le phènomène de prise de poids est homogène au sein de la population étudiée et qu'il est indépendant du poids initial des individus.

IV - (Session de Juin 2015 - Extrait)

Dans le cadre d'un enseignement d'informatique, chacun des étudiants de deux filières, (Master d'économie et Master de Droit) est invité à remettre un sujet de projet avant les vacances de Noël. On sait qu'habituellement la moitié d'entre eux remet un sujet à cette date.

1. Quelle est la probabilité que plus des $\frac{2}{3}$ des étudiants de Master d'économie remettent un sujet, s'ils sont au nombre de 49?
2. Quelle est la probabilité que l'écart entre les deux taux de remise des deux groupes d'étudiants excède 17 unités de %, les étudiants de Master de Droit étant au nombre de 169?

(On donne : $\sqrt{169} = 13$; $\sqrt{218} \simeq 15$)

V - (Session de Juin 2016)

Sur la planète Statistika, deux espèces intelligentes cohabitent : les "Simples" et les "Doubles". En notant respectivement X^S et X^D les tailles des individus (en cm) des races "Simples" et "Doubles", on sait que $X^S = LN(m, \sigma)$ et $X^D = LN(2m, 2\sigma)$. D'autre part, dans la population totale de la planète, les "Simples" sont en proportion p.

1. On note X la variable aléatoire représentant la taille d'un individu quelconque de la planète Statistika.
   - a. Donner l'expression de X en fonction de X^S et X^D.
   - b. Quelle est la loi suivie par la variable aléatoire X ? Justifier votre réponse.
   - c. Calculer $E[X]$ l'espérance de X et $Var[X]$ la variance de X en fonction des paramètres m, σ et p.
   - d. On prélève au hasard un individu de chaque race. A partir de quelle valeur de ($\frac{m}{\sigma}$) la probabilité que $\{X^D \geq X^S\}$ devient-elle supérieure à 0,8413 ?
2. On pose $p = \frac{2}{3}$. On prélève au hasard un échantillon de 100 individus de la planète et on note F_{100} la variable aléatoire représentant la proportion de "Simples" observée dans cet échantillon.
   - a. Déterminer complètement la loi de probabilité suivie par cette variable aléatoire. Justifier soigneusement votre réponse (on donne $\sqrt{2} = 1,4$).
   - b. Calculer $\alpha = prob\{0,6 \leq F_{100} \leq 0,7\}$.
   - c. Construire et calculer un intervalle de fluctuation de niveau 0,95 pour la variable aléatoire F_{100}. On note $[a \quad b]$ cet intervalle.
   - d. En notant F_n la variable aléatoire représentant la proportion de "Simples" observée dans un échantillon aléatoire de taille n, quelle devrait être la taille minimale n, si l'on souhaite que $prob\{a \leq F_n \leq b\} \geq 0,9973.$ Commenter le résultat obtenu.
   - e. Dans l'échantillon initial de taille $n = 100$, on observe 64 "Simples" et 36 "Doubles". Ces valeurs sont-elles compatibles avec les résultats obtenus aux questions **2**.**b**. et **2**.**c**.. Justifier votre réponse.
3. On pose maintenant $m = 110\,cm$ et $\sigma = 6\,cm$. On note respectivement $\bar{X}^S$ et $\bar{X}^D$ les variables aléatoires représentant les tailles moyennes observées dans les deux sous-échantillons de "Simples" et de "Doubles".
   - a. Déterminer précisément les lois de probabilité suivies par ces deux variables aléatoires.
   - b. Calculer $\beta = prob\{\bar{X}^D \geq 2\bar{X}^S\}$. Ce résultat était-il prévisible ?
   - c. En notant $\bar{X}_{100}$ la variable aléatoire représentant la taille moyenne des individus dans l'échantillon initial, donner l'expression de $\bar{X}_{100}$ en fonction de $\bar{X}^S$ et de $\bar{X}^D$.
   - d. Déduire de ce qui précède la loi suivie par la variable aléatoire $\bar{X}_{100}$.
   - e. Calculer $E[\bar{X}_{100}]$ l'espérance de $\bar{X}_{100}$ et $Var[\bar{X}_{100}]$ la variance de $\bar{X}_{100}$.

VI - Considérons deux populations de grande taille (population A et population B). Dans la population A, la proportion d'individus possédant un certain caractère qualitatif est $p_A = \frac{1}{5}$ tandis que dans la population B cette proportion est $p_B = \frac{2}{5}$. On prélève dans chaque population un échantillon aléatoire de taille $n = 100$. On note respectivement F_n^A et F_n^B les proportions d'individus possèdant le caractère étudié dans les échantillons issus des populations A et B.

1. Déterminer complètement la loi de probabilité suivie par la variable aléatoire F_n^A (Toute la démarche permettant de parvenir au résultat sera détaillée). En déduire celle suivie par la variable aléatoire F_n^B.
2. Calculer $\alpha_1 = prob\{0,2 \leq F_n^A \leq 0,3\}$ et $\alpha_2 = prob\{0,2 \leq F_n^B \leq 0,3\}$. Ce résultat était-il prévisible et pourquoi ?
3. Construire et calculer un intervalle de fluctuation de niveau $0,99$ pour la variable aléatoire F_n^A.
4. Calculer $\beta = prob\{F_n^A \geq F_n^B\}$. Pour calculer β, vous pourrez considérer la variable aléatoire $\bar{D} = F_n^B - F_n^A$ et déterminer complètement sa loi.
5. Quelle devrait être la taille des deux échantillons si l'on veut que la probabilité précédente β soit au moins égale à $0,05$. Justifier soigneusement votre réponse.
6. Expliquer ce dernier résultat.

NB : On donne $\sqrt{10} \simeq 3,2$; $\sqrt{6} \simeq 2,5$

VII - (**Session de Juin 2014**)

On note X la variable aléatoire représentant l'accroissement annuel de taille (en cm) des enfants d'une certaine classe d'age. Dans la population, on sait que $E[X] = m_x = 6,5$ et que $Var[X] = \sigma_X^2 = 16$.

1. Expliquez à l'aide d'un calcul simple pourquoi **il n'est pas raisonnable** de supposer que X suit une loi de Laplace-Gauss.
2. On prélève parmi les individus de cette classe d'âge, un échantillon aléatoire de 400 enfants, et on observe l'évolution de leur taille pendant un an. On note $\bar{X}_{400} = \frac{1}{400}\sum_{i=1}^{400} X_i$.
   - **a**. Que représente la variable aléatoire $\bar{X}_{400}$?
   - **b**. Déterminer complètement la loi de probabilité de $\bar{X}_{400}$. Justifier votre réponse de la façon la plus précise possible.
   - **c**. Construire un intervalle ayant 92% de chances de contenir $\bar{X}_{400}$.
   - **d**. Quelle doit être la taille n de l'échantillon à prélever si l'on souhaite avoir au moins 99,5% de chances d'observer la moyenne d'échantillonnage dans l'intervalle précédent ?
3. Dans la population étudiée, la croissance de la taille n'est pas homogène car, au sein de cette classe d'âge, les filles grandissent plus que les garçons. En effet, chez les garçons $E[X] = m_x^G = 5,5$ alors que chez les filles $E[X] = m_x^F = 8$. Par contre la dispersion est la même dans les deux groupes et est égale à celle observée pour l'ensemble de la population ($\sigma_{X,G}^2 = \sigma_{X,F}^2 = \sigma_X^2 = 16$).
   - **a**. Déduire de ce qui précède, la part de la population masculine au sein de la classe d'âge étudiée.
   - **b**. L'échantillon initialement prélevé comprend 256 garçons et 144 filles. On notera $\bar{X}_G$ et $\bar{X}_F$ les moyennes d'échantillonnages observées sur les 2 sous-échantillons.
     - **i**. Déterminer complètement les lois suivies par les variables aléatoires $\bar{X}_G$ et $\bar{X}_F$.
     - **ii**. Calculer $\alpha = prob\{\bar{X}_G \geq \bar{X}_F\}$. Le résultat obtenu vous parait-il logique ?
     - **iii**. Pour quelle valeur de m_x^G observerait-on $\alpha = \frac{1}{2}$? Justifier votre réponse de façon précise.

c. On note F_n la variable aléatoire représentant la proportion de garçons dans un échantillon de taille n prélevé aléatoirement au sein de la classe d'âge étudiée.

i. Déterminer la loi suivie par la variable aléatoire F_n en supposant que n est suffisamment grand.

ii. Construire un intervalle de fluctuation de niveau 0,9973 pour la variable F_n.

iii. En posant $n = 400$, déterminer numériquement les bornes de cet intervalle et confronter ce résultat avec la valeur de F_n effectivement observée dans l'échantillon de 400 enfants.

VIII - (**Session de Juin 2018 - Extrait**)

Dans un garage de réparation automobile, depuis le début de l'année 2018, le montant moyen d'une facture est de 75 euros pour un écart-type de 90 euros. On notera X la variable aléatoire représentant le montant d'une facture quelconque.

1. Ces données permettent-elles de penser que la distribution de la variable X est gaussienne ? Justifier votre réponse.

2. Pour les 225 factures émises lors de la dernière semaine d'avril, le montant moyen est de 85 euros.On note $\bar{X}$ le montant moyen observé sur un échantillon aléatoire de taille $n = 225$.

a. Déterminer complètement la loi de la variable aléatoire $\bar{X}$, en justifiant soigneusement votre réponse.

b. Construire un intervalle de fluctuation de niveau 0,985 pour la variable aléatoire $\bar{X}$. On notera $[a\,,\,b]$ cet intervalle.

c. Calculer la probabilité qu'un échantillon aléatoire de cette taille extrait de la population habituelle donne un montant moyen au moins aussi écarté de la valeur attendue que l'est 85.

d. En notant $\bar{X}_n$ la variable aléatoire représentant le montant moyen calculé sur un échantillon aléatoire de n factures, à partir de quelle taille d'échantillon doit-on observer : $prob\,\{a \le \bar{X}_n \le b\} \ge 0,95$?

e. Si une autre semaine donne lieu à l'émission d'un nombre de factures moins élevé que 225, le garage a-t-il plus ou moins de chances que pour la dernière semaine d'avril de trouver un montant moyen supérieur à 85 euros. La solution de cette question doit être obtenue **sans calcul**, **seulement au moyen d'un raisonnement**.

3. Les trois quarts des factures sont des factures de mécanique et le quart restant représente des factures de carrosserie. On note F_1 la variable aléatoire représentant la proportion de factures de mécanique observée sur un échantillon aléatoire de taille $n = 225$.

a. Déterminer complètement la loi de la variable aléatoire F_1, en justifiant soigneusement votre réponse.

b. Dans le lot de 225 factures de la dernière semaine d'avril, 150 seulement sont des factures de mécanique. Cette proportion observée peut-elle apparaitre comme significativement différente de la proportion attendue ? On considérera la différence comme significative si elle a moins de 1 chances sur 100 de se produire.(On donne $\sqrt{3} = 1,7$)

c. La première semaine de mai a donné lieu à l'établissement de k factures parmis lesquelles on a pû décompter 130 factures de mécanique. Notons F_2 la variable aléatoire représentant la proportion de factures de mécanique observée sur un échantillon aléatoire de taille k. Sachant que $prob\left\{F_2 \ge \frac{130}{k}\right\} = 0,281$, en déduire la valeur de k, nombre de factures établies par le garage lors de la première semaine de mai. On se contentera ici de trouver une équation sans la résoudre.

IX - (Session de Mai 2017 - Extrait)

Une machine conditionne des pruneaux en poches de 25 unités. On sait que le poids d'un pruneau, noté X, est une variable aléatoire qui suit une loi normale d'espérance $m = 10$ grammes et d'écart-type $\sigma = 1$ gramme. On note P la variable aléatoire représentant le poids en pruneaux d'une poche quelconque.

1. Quelle est la loi suivie par P ? Justifier votre réponse.
2. Calculer les paramètres de la loi de P (son espérance notée $E(P)$ et son écart-type noté σ_P).
3. Les normes Européennes imposent d'indiquer sur la poche le poids net des pruneaux conditionnés. On décide de fournir ce renseignement sous la forme suivante :

$$Poids \quad Net \; : \; 250\, grammes \pm \alpha\, grammes$$

   a. Déterminer α de telle sorte que l'on ait 97 % de chances de trouver le poids en pruneaux d'une poche quelconque à l'intérieur de cet intervalle.
   b. La valeur précédente n'étant pas entière, on décide pour simplifier de choisir $\alpha = 10$ grammes. Calculer la probabilité (notée β) qu'en prélevant alors une poche au hasard dans la production on trouve un poids de pruneaux dans l'intervalle correspondant.
4. Un futur partenaire en charge de la commercialisation des poches de pruneaux décide de vérifier la pertinence des renseignements inscrits sur l'emballage. Il prélève pour ce faire, trois poches au hasard dans la production de la période qui en contient beaucoup, et pèse les pruneaux contenus dans chacunes d'entre elles. Il trouve pour deux des trois poches, un poids de pruneaux à l'extérieur de l'intervalle mentionné (celui de la question **3-b**). Evaluer la probabilité d'un tel évènement en justifiant soigneusement votre réponse.

X- Problème de révision - (Session de Juin 2017)

Après correction des copies, la distribution des notes d'un examen national impliquant un grand nombre de candidats peut être considérée comme une variable aléatoire (notée X) d'espérance égale à 9 et d'écart-type égal à 3.

1. A l'aide de l'inégalité de Bienaymé-Tchebychev, calculer un majorant de la probabilité que X diffère de son espérance de plus de 2 fois son écart-type. Vous fournirez une bonne raison d'utiliser cette inégalité dans le cas présent.
2. On prélève au hasard un échantillon de n copies corrigées et on note $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la variable aléatoire égale à la moyenne des notes de cet échantillon.
   a. Donner la loi de $\bar{X}_n$ (justifier soigneusement les hypothèses), son espérance et son écart-type.
   b. Construire un intervalle de fluctuation de niveau 0,975 pour la variable $\bar{X}_n$.
3. Pour la saisie administrative des notes, chaque employé du centre d'examen effectue son travail par tas de n copies. Deux employés du centre d'examen travaillent en même temps et saisissent les notes d'un tas de n copies chacun. Soit α la probabilité que la moyenne (sur les n copies) obtenue par l'un des employés soit supérieure d'au moins 0,5 point à celle obtenue par l'autre.
   a. Evaluer cette probabilité en fonction de la valeur de n.
   b. Représenter graphiquement l'évolution de α en fonction de n. Ce résultat était-il prévisible ? Pourquoi ?
   c. A partir de quelle valeur de n cette probabilité devient elle inférieure à 2,5% ?
4. On considère maintenant deux correcteurs distincts de cet examen. Les deux correcteurs ne sont pas aussi indulgents l'un que l'autre. On peut considérer que les notes données par les deux correcteurs (notés respectivement 1 et 2) ont les caractéristiques suivantes :

Correcteur n°1 : $E[X_1] = 9 + \mu$ $\sigma_1 = 3 - \mu$

Correcteur n°2 : $E[X_2] = 9 - \mu$ $\sigma_2 = 3 + \mu$

où $\mu \in [0\,;3]$. On prélève aléatoirement deux échantillons de 100 copies corrigées par chacun des deux correcteurs et on note $\bar{X}_1$ et $\bar{X}_2$ les variables aléatoires correspondant aux moyennes respectives obtenues dans les deux échantillons.

a. Déterminer complètement les lois suivies par les variables $\bar{X}_1$ et $\bar{X}_2$.

b. Représenter shématiquement les fonctions de densités de ces deux variables sur un même graphique.

c. On note β la probabilité de l'évènement $\{\bar{X}_1 \geq \bar{X}_2\}$.

i. En notant $\bar{D} = \bar{X}_1 - \bar{X}_2$, déterminer complètement la loi de $\bar{D}$.

ii. Calculer β en fonction de μ.

d. Pour quelle valeur de μ cette probabilité est-elle minimale ? Que vaut β dans ce cas là ? Expliquer ce résultat.

e. Calculer la valeur de μ associée à $\beta = 0,975$.

5. On suppose maintenant que les distributions des notes des deux correcteurs sont gausiennes et on s'intéresse au cas d'un étudiant ayant absolument besoin d'une note supérieure ou égale à 15 sur 20 à cet examen.

a. En quoi les hypothèses sur les lois de X_1 et X_2 sont elles nécessaires dans cette situation ?

b. Quel correcteur préfèrera-t'il se voir désigner, en fonction de la valeur prise par le paramètre μ ? (ici $\mu \in [0\,;3[$).

XI - Problème de révision - (Session de Mai 2018)

Après correction des copies, la distribution des notes de l'épreuve de statistiques (notée sur 20 points) d'un concours national impliquant un grand nombre de candidats fait apparaître une moyenne $m = 8,5$ pour un écart-type $\sigma = 2$.

1. En notant X la variable aléatoire représentant la note obtenue par un candidat quelconque, évaluer $\alpha = prob\{2,5 \leq X \leq 14,5\}$. Expliquer et justifier précisément la méthode utilisée.

2. On prélève, dans l'ensemble des copies, un échantillon représentatif de 36 copies.

a. Calculer la probabilité que la note moyenne de ces 36 copies soit comprise entre 8 et 10. Toute la démarche permettant de parvenir au résultat sera détaillée.

b. Construire un intervalle de fluctuation de niveau 0,93 pour la note moyenne de ces 36 copies. On notera $[a, b]$ cet intervalle.

c. Quelle taille minimale d'échantillon devrions nous prélever si l'on souhaitais que la probabilité que la note moyenne observée appartienne à l'intervalle $[a, b]$ dépasse 99,5% ?

3. On prélève dans l'ensemble des copies un second échantillon aléatoire de 64 copies. Calculer la probabilité que l'écart entre les notes moyennes des deux échantillons soit supérieur à 1,2 point.

4. Pour le centre d'examen de Bordeaux, 100 candidats ont passé l'épreuve de statistiques. La note moyenne observée dans ce centre est égale à 10,2. Peut-on considérer ce groupe d'étudiants comme représentatif de l'ensemble des étudiants ayant passé l'épreuve ? Justifier soigneusement votre réponse.

5. Parmi les 100 candidats bordelais on compte 36 filles. La note moyenne de ces candidates est de 11 sur 20.

a. Déduire de ce qui précède la note moyenne des candidats masculins du centre d'examen de Bordeaux.

b. La différence des notes moyennes, observée entre les deux groupes, est-elle significative ? Justifier votre réponse.