LICENCE Sciences Economiques 2^{ième} année - Statistiques II

Année Universitaire 2012 - 2013

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 - Un sondage a été réalisé dans deux communes, A et B. Dans chacune ont été désignés au hasard un certain nombres d'habitants auxquels la même question a été posée. On précise les données suivantes :

Commune	Taille de la population	Taille de l'échantillon	Taux de sondage
A	2000	200	10 %
B	50000	400	0,8 %

1) Dans quelle commune les résultats (réponses à la question) sont-ils les plus précis ? (Préciser la notion statistique de précision).
2) Déterminer le rapport existant entre les deux mesures de cette précision.
3) Que devient cette précision lorsque les échantillons ont pour tailles réspectives,
   a) 2000 dans chaque commune ?
   b) 10000 dans chaque commune ?

IV - 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.

V - Un producteur de poires remplit des cagettes préformées avec 25 poires calibrées, poires dont on sait que le poids moyen unitaire est de 120 grammes pour un écart-type de 12 grammes.

1) Quelle est la probabilité qu'une cagette ait un poids de poires ne dépassant pas 2,9 kg ?
2) Quelles sont les limites de confiance à 95 % pour le poids de poires dans une cagette ?
3) Quelle est la probabilité que la différence de poids de poires entre deux cagettes excède la moitié du poids moyen d'une poire ?

VI - Au second tour d'une élection nationale, un des deux candidats, le candidat A, obtient 6 695 000 voix sur les 13 000 000 de suffrages exprimés.

1) La veille du vote, un institut de sondage annonçait la victoire du candidat B sur la base du résultat obtenu sur un échantillon représentatif de 100 personnes. Evaluer la probabilité qu'avait cet institut de se tromper. Toute la démarche permettant de parvenir au résultat sera développée.
2) Quelle aurait été cette même probabilité si l'institut avait prélevé un échantillon représentatif de 1600 personnes ?
3) Comment peut-on expliquer ce résultat ?

VII - 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 confiance 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 $\sqrt{10} \simeq 3,2$; $\sqrt{6} \simeq 2,5$

VIII - Un sujet d'examen proposé aux nombreux étudiants d'une université scientifique comprend trois exercices, notés respectivement sur 4, 6 et 10 points (pour un total de 20 points), portant chacun sur une des trois parties du cours. Sachant que chaque étudiant connaît très bien une partie du cours (il aura le maximum à l'exercice relatif à cette partie), moyennement une seconde (il aura la moitié des points à l'exercice relatif à cette partie) et peu la troisième (il n'aura aucun point à l'exercice relatif à cette partie) et que, pour un étudiant, le choix des parties étudiées a été fait au hasard, calculer une valeur approchée de la probabilité

1) qu'un groupe aléatoire de trente étudiants ait une note moyenne supérieure à onze points ?
2) que la différence entre les notes moyennes de deux groupes aléatoires de trente étudiants excède 1,5 point ?

IX - (Session de Mai 2007)

Considérons une population A , de grande taille. On note X la variable aléatoire représentant la taille (exprimée en cm) d'un individu quelconque. On note $E[X] = m_A$ et $Var[X] = \sigma_A^2$.
On prélève un échantillon aléatoire de 100 individus dans cette population. On note $\bar{X}_{100}^A$ la taille moyenne des individus de l'échantillon.

1) Expliquer pourquoi on peut admettre que $\bar{X}_{100}^A$ suit une loi normale.
2) Peut-on en déduire que X suit aussi une loi normale ? Justifier votre réponse.
3) On donne $E[\bar{X}_{100}^A] = 175$ et $\sigma_{\bar{X}_{100}^A} = 5$. En déduire les valeurs prises par $E[X]$ et $Var[X]$. Justifier vos réponses.
4) Construire et calculer un intervalle de confiance de niveau 0,9973 pour la variable aléatoire $\bar{X}_{100}^A$. On notera a et b $(a < b)$ les deux bornes de cet intervalle.

5) En notant $\bar{X}_n^A$ la taille moyenne des individus d'un échantillon aléatoire de taille n prélevé dans la même population A, calculer la valeur de n telle que : $prob\{a < \bar{X}_n^A < b\} = 0,95.$ Ce résultat vous semble t'il logique ? Argumenter votre réponse de façon aussi détaillée que possible.
6) Considérons une seconde population (notée B), elle aussi de grande taille. En notant toujours X la variable aléatoire représentant la taille d'un individu quelconque de cette population B, on a $E[X] = m_B = 170$ et $Var[X] = \sigma_B^2$. On prélève dans cette population un échantillon aléatoire de taille $n = 100.$ On notera $\bar{X}_{100}^B$ la taille moyenne des individus de cet échantillon.
   a) Déterminer complètement la loi suivie par $\bar{X}_{100}^B$.
   b) Sachant que $prob\{\bar{X}_{100}^A > \bar{X}_{100}^B\} = 0,6915$, calculer la valeur de σ_B (écart-type des tailles au sein de la population B) compatible avec cette valeur de la probabilité. (On donne $\sqrt{3} = 1,7$).
   c) Quelle est la valeur maximale que peut prendre cette dernière probabilité ?

X - (Session de Juin 2007)

Considérons une population de grande taille. On sait que dans cette population, la proportion des individus possèdant une connection internet haut-débit à leur domicile est de $\frac{3}{4}$.

On prélève dans cette population un échantillon aléatoire de 300 individus. Soit Q_1 la variable aléatoire représentant le nombre d'individus de cet échantillon possèdant une connection internet haut-débit à leur domicile.

1) Quelle est la loi suivie exactement par la variable aléatoire Q_1 ? Justifier soigneusement votre réponse.
2) Calculer le plus simplement possible $E(Q_1)$ et $\sigma(Q_1)$ son espérance et son écart-type.
3) Calculer $\alpha_1 = prob\{215 < Q_1 < 240\}.$ Présenter complètement la démarche permettant de parvenir au résultat.
4) On sait que la population étudiée n'est pas homogène du point de vue de la possession ou non d'une connection internet haut-débit. Plus précisément, 90% des urbains en possèdent une tandis que seulement 40% des ruraux sont dans ce cas. Déduire de ce qui précède, la part de la population urbaine dans cette population.
5) On souhaite tenir compte de cette non-homogénéité de la population lors du prélèvement de l'échantillon de taille 300.
   a) Nommer et présenter de façon concrète la méthode permettant d'y parvenir.
   b) En notant Q_2 la variable aléatoire représentant le nombre d'individus de ce second échantillon possèdant une connection internet haut-débit à leur domicile, montrer que Q_2 suit approximativement une loi normale.
   c) Calculer $E(Q_2)$ et $\sigma(Q_2)$ l'espérance et l'écart-type de cette variable aléatoire.
   d) Soit $\alpha_2 = prob\{215 < Q_2 < 240\}$. Déduire de ce qui précède, et sans calcul, que $\alpha_2 > \alpha_1$. Commenter ce dernier résultat.

XI - (Session de Juin 2009)
Soit X un caractère quantitatif distribué normalement au sein d'une population de grande taille. On sait que $E[X] = m_X = 100$ et que $Var(X) = \sigma^2 = 100$. On prélève au hasard dans cette population un échantillon de taille n_1. On note $\bar{X}_1$ la moyenne d'échantillonnage observée dans cet échantillon.
On prélève à nouveau un échantillon aléatoire dans la même population, échantillon de taille $n_2 = k.n_1$ avec $k > 1$. On note $\bar{X}_2$ la moyenne d'échantillonnage observée dans ce second échantillon.

1) Donner les expressions des variables aléatoires $\bar{X}_1$ et $\bar{X}_2$.
2) Déterminer complètement les lois de probabilité suivies par $\bar{X}_1$ et $\bar{X}_2$.
3) Soit $\bar{D}$ la variable aléatoire définie par : $\bar{D} = \bar{X}_1 - \bar{X}_2$.
   a) Quelle est la loi suivie par $\bar{D}$? Justifier votre réponse.
   b) Calculer $E[\bar{D}]$, l'espérance mathématique de $\bar{D}$ et $\sigma_{\bar{D}}$ l'écart-type de $\bar{D}$ en fonction de k et de n_1.
   c) Expliquer en une phrase pourquoi on peut savoir sans calcul que $prob\{\bar{X}_2 \geq \bar{X}_1\} = \frac{1}{2}$.
   d) Sachant que $prob\left\{\bar{D} \geq 1\right\} = 0,1587$, déterminer la relation liant les paramètres k et n_1.
   e) Déterminer la valeur de n_1 telle que $prob\left\{\bar{X}_1 \geq 102\right\} = 0,0125$. (Arrondir n_1 à l'entier le plus proche).
   f) Déduire des questions précédentes la valeur de n_2, taille du second échantillon.
   g) Calculer $\beta = prob\left\{\bar{X}_2 \geq 102\right\}$. Expliquer le résultat trouvé.

XII - Problème de révision - (Session de Mai 2004)

Les parties 1 et 2 sont indépendantes
PARTIE 1 - On se propose de simuler des réalisations d'une variable aléatoire X, laquelle suit une loi normale d'espérance mathématique 8,5 et d'écart-type 2,5.

1) Exposer rapidement le principe de la méthode de l'anamorphose permettant de réaliser une telle simulation. Donner une illustration graphique de cette méthode.
2) On fournit l'extrait suivant d'une table de nombres au hasard :

$$9861 \quad 1587 \quad 0228 \quad 9641 \quad 3446 \quad 0808 \quad 9452 \quad 0548 \quad 9953$$

   a) En considérant des groupes de 4 chiffres dans la liste précédente et en utilisant la méthode présentée à la question 1, réaliser la simulation de 9 valeurs particulières de la variable aléatoire X.
   b) Calculer $\bar{X}_9$ la moyenne d'échantillon et S^2 la variance estimée. On donne

$\sum_{i=1}^{9}(X_i - \overline{X}_9)^2 = 167$. La quantité $\sum_{i=1}^{9}(X_i - \overline{X}_9)^2$ est appelée Somme des Carrés des Ecarts (SCE) .

PARTIE 2 - Après correction des copies, la distribution des notes de l'épreuve de statistique (notée sur 20 points) d'un examen national impliquant un grand nombre de candidats, suit approximativement une loi normale d'espérance mathématique $m = 8,5$ et d'écart-type $\sigma = 2,5$. On prélève aléatoirement 9 copies dans l'ensemble des copies corrigées. On appellera X_i la note donnée à la $i^{ème}$ copie.

1) En notant $\overline{X}_9$ la moyenne des notes de cet échantillon aléatoire, déterminer complètement la loi de probabilité de $\overline{X}_9$. Justifier votre réponse.
2) Construire et calculer un intervalle de confiance centré de niveau $\alpha = 0,95$ pour la moyenne d'échantillonnage $\overline{X}_9$.
3) On procède au tirage aléatoire d'un second échantillon de 9 copies dans l'ensemble des copies corrigées. Calculer la probabilité que l'écart de notes moyennes entre les deux échantillons excède 2 points. On donne $\sqrt{2} = 1,4$.
4) Quelle devrait être la taille minimale commune des deux échantillons si l'on désire que la probabilité précédente ne dépasse pas 1% ? Commentez.

PARTIE 3 - On considère maintenant que les 9 valeurs obtenues au cours de la simulation de la première partie sont en fait les notes associées à un échantillon fictif de 9 copies du même examen national.

1) La valeur de $\overline{X}_9$ calculée au cours de la première partie est-elle compatible avec l'intervalle de confiance trouvé dans la seconde partie ? Celà permet-il de considérer cet échantillon comme représentatif de l'ensemble de la population concernée ? Justifier votre réponse.
2) On suppose que, pour être reçu à l'examen, il faut obtenir une note supérieure ou égale à 10 sur 20, qu'une note inférieure ou égale à 6 sur 20 se traduit par l'élimination définitive du candidat et que dans tous les autres cas, le candidat est authorisé à passer une épreuve de rattrapage.
   a) On prélève au hasard, dans l'ensemble des copies, la copie d'un des candidats à l'examen. Calculer la probabilité que ce candidat passe l'épreuve de rattrapage ?
   b) Comparer cette valeur avec la distribution des fréquences relatives correspondante dans l'échantillon fictif de la première partie. Que vous suggère cette comparaison ?
3) On procède au tirage aléatoire d'un troisième échantillon de 9 copies dans l'ensemble des copies corrigées. Calculer la probabilité que la Somme des Carrés des Ecarts calculée sur cet échantillon soit supérieure ou égale à 167 .
4) En utilisant les résultats des 3 parties du problème, que pouvez-vous conclure quant à la

représentativité de l'échantillon fictif obtenu ? Justifier votre réponse et donner une explication plausible du problème rencontré.

XIII - Problème de révision - (Session de Mai 2012) - Correction en cours

Un élevage industriel de volailles, élève et commercialise des poulets, poulets dont on sait que le poids au moment de l'abattage (noté X) est une variable aléatoire normale de moyenne $m = 1500g$ et d'écart-type $\sigma = 100g$.

L'élevage commercialise les volailles, juste après l'abattage, par cartons de 25 unités.

1) En notant P_1 le poids en poulets d'un carton quelconque, donner l'expression de P_1 en fonction des variables X.
2) Déterminer complètement la loi de P_1. Justifier votre réponse.
3) Calculer les probabilités que P_1 soit supérieur à $40kg$ et que P_1 soit inférieur à $36,5kg$.
4) Construire et calculer un intervalle de confiance de niveau $0,985$ pour P_1 le poids en poulets d'un carton quelconque. On note $[a,b]$ cet intervalle.
5) Pour quelle valeur de σ pourrait-on obtenir le résultat suivant : $prob\{a \leq P_1 \leq b\} = 0,9973$?
6) L'élevage propose à ses clients deux formules d'achat : le client à le choix entre acheter le carton pour une somme forfaitaire de 225 Euros ou bien payer le poids réel des poulets du carton au prix de 6 Euros par kilo. Quelle solution est la plus économique a priori pour le client ? Justifier votre réponse.
7) Des études sérieuses ont montré que la perte de poids liée au plumage / vidage des poulets est une variable aléatoire Y qui suit une loi uniforme sur l'intervalle $[150g, 270g]$ et est indépendante du poids à vif des poulets. On notera Z la variable aléatoire représentant le poids d'un poulet une fois plumé et vidé.
   a) En utilisant des résultats donnés en cours, calculer $E[Y]$ l'espérance de Y et $Var[Y]$ la variance de Y
   b) En déduire $E[Z]$ l'espérance de Z et σ_Z l'écart-type de Z (on donne $\sqrt{112} = 10.6$).
   c) Que peut-on dire de la loi de probabilité suivie par la variable aléatoire Z ?
   d) Evaluer un minorant de la probabilité, qu'en prélevant au hasard dans la production un poulet plumé et vidé, son poids soit compris entre $1150g$ et $1430g$.
8) Les poulets plumés et vidés sont aussi commercialisés par cartons de 25 unités. On note P_2 la variable aléatoire représentant le poids en poulets d'un carton de ce type.
   a) Quelle est la loi de probabilité suivie approximativement par P_2 ? Justifier votre réponse.
   b) Déterminer $E[P_2]$ l'espérance de P_2 et $Var[P_2]$ la variance de P_2.

c) On note toujours P_1 la variable aléatoire représentant le poids en poulets d'un carton de 25 poulets justes abattus et P_3 la variable aléatoire représentant le poids en poulets d'un carton de n poulets plumés et vidés. A partir de quelle valeur de n avons-nous au moins 95% de chances d'observer $P_3 > P_1$? On se contentera ici de trouver une équation sans la résoudre.

XIV - Problème de révision - (Session de Mai 2008) - Correction en cours

Ayant besoin d'un grand nombre d'axes de 200 millimètres de diamètre, un industriel s'adresse à un fabricant. Ce dernier décide, avec son ingénieur de fabrication d'utiliser pour cette commande une machine pour laquelle on sait que la distribution des diamètres des axes usinés est une variable aléatoire Gaussienne, d'espérance égale à m et d'écart-type, noté σ, égal à 2 millimètres. On note X le diamètre en millimètres d'un axe usiné.

1) Au démarrage de la production, l'ingénieur fixe le réglage de la machine sur $m = 200$. Après quelques heures de production, il prélève un échantillon aléatoire de taille $n = 9$.
   a) Expliquer pourquoi, dans ce cas, une faible taille d'échantillon n'est pas génante.
   b) En notant $\bar{X}_9$ le diamètre moyen observé dans l'échantillon en question, déterminer complètement la loi suivie par cette variable aléatoire. Justifier votre réponse.
   c) Calculer $\alpha = prob\{199 \leq \bar{X}_9 \leq 201\}$.
   d) Calculer $[a,b]$ intervalle de confiance de niveau 0,985 pour la variable $\bar{X}_9$.
   e) Quelle devrait être la taille de l'échantillon si l'on souhaitait que la probabilité que la moyenne d'échantillonnage appartienne à l'intervalle précédent soit d'au moins 0,999.
2) La machine continue de produire dans les conditions décrites à la question précédente.
   a) En prélevant un nouvel échantillon aléatoire de 9 pièces dans la production, l'ingénieur touve la moyenne d'échantillonnage en dehors de l'intervalle $[a,b]$. Expliquer pourquoi il est rationnel que l'ingénieur pense alors que la machine se soit déréglée. On admettra par la suite, qu'il vérifie le réglage de la machine dès qu'il rencontre cette situation.
   b) Quelle est la probabilité qu'il se trompe ?
   c) Quelle devrait être la taille de l'échantillon à prélever s'il désirait diviser par 15 cette dernière probabilité ?
3) La machine se dérègle à l'insu de l'ingénieur et se met à usiner des pièces d'un diamètre moyen m' égal à 201 millimètres, la dispersion de la fabrication n'étant pas modifiée. Quelle est la probabilité qu'en prélevant alors un échantillon aléatoire de 9 pièces, l'ingénieur intervienne sur le réglage de la machine ?
4) La machine ayant travaillé longtemps, l'ingénieur décide de vérifier la stabilité de la variance de la fabrication. Un échantillon de 9 pièces, prélevées aléatoirement à cet effet, a fourni : $\sum_i (X_i - \bar{X}_n)^2 = 44$, où X_i est le diamètre d'une pièce fabriquée, i. On appelle **SCE** (Somme des Carrés des Ecarts à la moyenne) cette quantité.
   a) En déduire la valeur de S^2, variance estimée de la fabrication au moment du prélèvement de cet échantillon.
   b) (Hors barème) Si la dispersion n'a pas changé, quelle est la probabilité que S^2 soit supérieure à la valeur trouvée ?
   c) (Hors barème) Quelle conclusion l'ingénieur peut-il tirer du résultat précédent ?