FSEG STRASBOURG
Faculté des sciences économiques et de gestion

UNIVERSITÉ DE STRASBOURG

Année universitaire 2013/2014

LICENCE 1ère année Économie-Gestion et Mathématiques-Économie

Semestre 1 – Session 2 / Contrôle terminal / Juin 2014

Probabilités-statistique I (Giuseppe Attanasi et Luc Naegele)

Durée : 1 heure 30

Tous documents interdits

Calculatrice autorisée

Barème de notation :

Exercice 1 : Questionnaire à choix multiples (5 points)

Pour chaque question :

Si la (les) réponse(s) est (sont) exacte(s)	+ 1 point
Si au moins une réponse est fausse	- 1 point
En cas de non réponse	0 point

Exercice 2 : Vrai ou faux ? (6 points)

Pour chaque proposition : 1 point.

Exercice 3 : Le revenu nécessaire pour vivre (9 points + 3 points bonus)

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Barème	0,5	0,5	0,5	0,5	1	0,5	0,5	2	(+1)	(+1)	1	1	1	(+1)

Les questions 9, 10 et 14 sont des questions subsidiaires : vous n'êtes pas obligés d'y répondre mais elles peuvent chacune rapporter un point supplémentaire.

Exercice 1 : Questionnaire à choix multiples

Indiquez la (ou les) bonne(s) réponse(s) sur votre copie. Aucune justification n'est demandée.

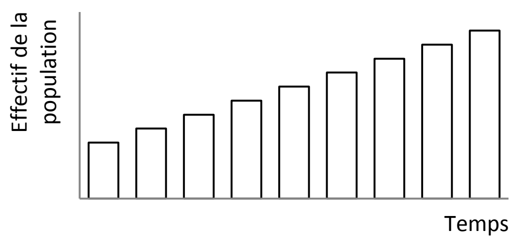
1. « Nous pouvons donc tenir pour certain que, lorsque la population n'est arrêtée par aucun obstacle, elle va doublant tous les vingt-cinq ans [...] » (T. R. Malthus, 1798, *Essai sur le principe de population*). Quel serait alors le taux de croissance annuel moyen de la population s'il n'y a aucun « obstacle » pour freiner sa progression ?
   a) Environ 2 %
   b) Environ 2,7 %
   c) Environ 2,8 %
   d) Environ 3 %
   e) 4 %
   f) Aucune réponse correcte

   On peut le vérifier à l'aide de la « règle des 70 » : $\frac{70}{25} = 2{,}8$

   Ou alors en calculant un accroissement annuel moyen : $\sqrt[25]{2} \approx 1{,}028$

2. La figure 1 représente l'évolution d'une population de période en période. Nous pouvons dire que le taux de croissance de cette population dans le temps...
   a) ... est constant
   b) ... est de plus en plus élevé
   c) ... est de plus en plus faible
   d) ... est stable
   e) ... est relativement faible
   f) ... est relativement élevé
   g) Aucune réponse correcte

   Figure 1

   De périodes en périodes, la variation absolue de l'effectif est constante : il y a x individus en plus à chaque période. Par conséquent le taux de croissance de la population, c'est-à-dire la variation de l'effectif pendant une période par rapport à l'effectif en début de période, est de plus en plus petit.

3. Soit une variable X qui prend n valeurs. On connaît la somme des carrés des valeurs de X et la valeur moyenne de X, respectivement :

$$\sum_{i=1}^{n} x_i^2 = 5000$$

$$\bar{X} = 6$$

   3.1. Si on considère que $n = 100$, alors que vaut la variance de X ?
   a) $\sigma_X^2 = 1$
   b) $\sigma_X^2 \approx 3{,}74$
   c) $\sigma_X^2 = 4$
   d) $\sigma_X^2 = 14$
   e) $\sigma_X^2 \approx 28{,}89$
   f) Aucune réponse correcte

   On utilise la relation de Koenig selon laquelle la variance est égale à la différence entre la moyenne des carrés et le carré de la moyenne : $\sigma_X^2 = \overline{X^2} - \bar{X}^2 = \frac{5\,000}{100} - 6^2 = 14$

3.2. Si on considère que l'écart-type de X est $\sigma_X \approx 5$, alors combien vaut n ?

a) $n = 82$

b) $n = 100$

c) $n = 118$

d) $n = 122$

e) $n = 180$

f) Aucune réponse correcte

On peut remarquer que $\sigma_X \approx 5 > \sqrt{14}$ ce qui implique que n est plus faible qu'à la question 3.1, ce qui ne nous laisse que la réponse a). Sinon, on calcule : $n = \frac{\sum_{i=1}^{n} x_i^2}{\sigma_X^2 + \bar{X}^2} \approx 82$.

4. Soit une série statistique. On peut dire que la moitié des valeurs de la série…

a) … sont supérieures ou égales à la médiane

b) … sont inférieures ou égales à la médiane

c) … sont supérieures ou égales au 5e décile

d) … sont relativement proches de la moyenne

e) … sont comprises entre le 1er et le 3e quartiles

f) … sont des valeurs extrêmes

g) Aucune réponse correcte

Ces quatre affirmations sont vraies par définition.

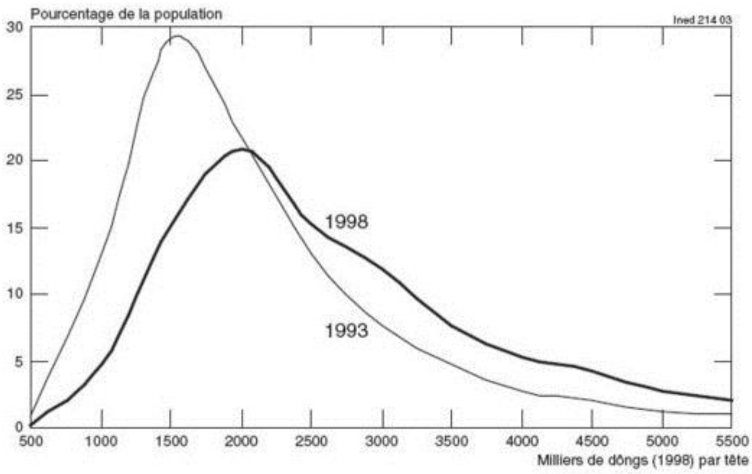
Exercice 2 : Vrai ou faux ?

Pour chaque proposition, indiquez « vrai » ou « faux » sur votre copie et expliquez votre choix. Attention : une réponse non justifiée ne rapporte aucun point.

1. Si les salaires moyens de chaque catégorie socioprofessionnelle salariée augmentent tous de 10 % durant une période, alors le salaire moyen augmente aussi de 10 %.

   Faux. Ce n'est vrai que si la structure de la population salariée reste la même au cours de cette période. Si les fréquences des différentes catégories dans la population totale varient alors le salaire moyen n'augmentera pas de 10 %. Toutefois la formulation de la question manque de précision et d'autres réponses ont été acceptées.

Figure 2 : Distribution des dépenses totales des ménages au Vietnam (en milliers de dôngs de 1998 par personne)



Note : Le dông est la monnaie vietnamienne.
Source des données : Enquête Niveau de vie au Vietnam.
Source du graphique : T. Minh Nguyen et B. M. Popkin, 2003, « Évolution des revenus et du système de santé au Vietnam : réduction de la pauvreté et augmentation des inégalités de prise en charge », *Population*, volume 58

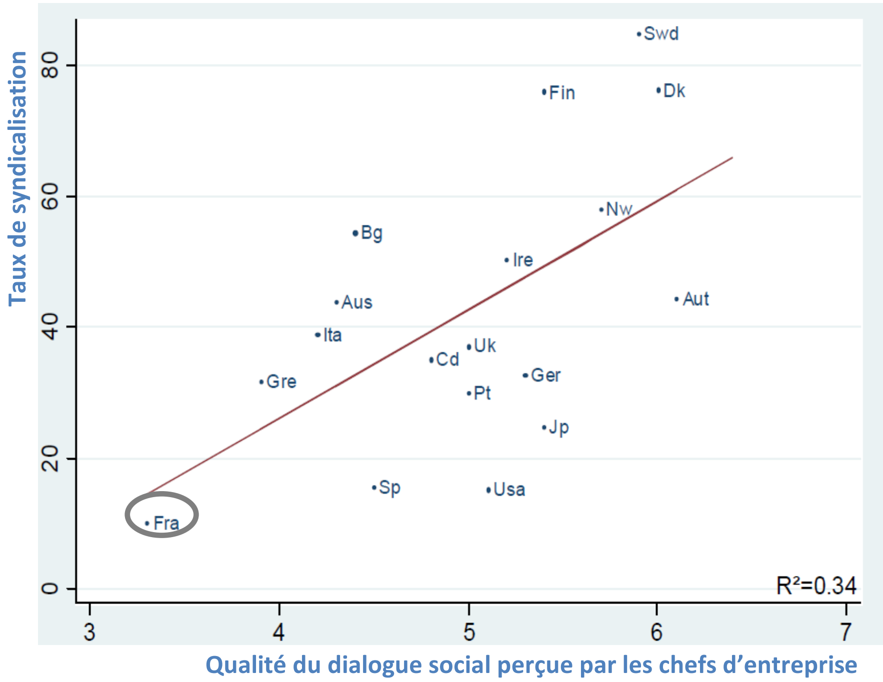
2. D'après la figure 2, la dépense moyenne des ménages vietnamiens a diminué entre 1993 et 1998.

   C'est faux. Les deux distributions représentées montrent que les ménages qui dépensent moins de 2 000 dôngs par personne et par an sont plus fréquents en 1993 qu'en 1998. À l'inverse, les dépenses de plus de 2 000 dôngs par tête sont plus fréquentes en 1998. Par exemple en 1993 plus de 10 % des ménages dépensaient 1 000 dôngs par an tandis qu'ils ne sont que 4 % environ à dépenser si peu en 1998. À noter que le titre de l'article dont est tiré le graphique nous fournit un indice : « réduction de la pauvreté ».

3. La figure 2 montre qu'en 1993 et en 1998 la dépense moyenne est plus élevée que la dépense médiane.

   C'est vrai. Les deux distributions sont étalées à droite, or les valeurs du haut de la distribution (les dépenses les plus élevées) influencent la moyenne mais pas la médiane. L'asymétrie de la distribution se traduit ici par une moyenne plus élevée que la médiane.

Figure 3 : Corrélation entre le taux de syndicalisation et la qualité du dialogue social perçue par les chefs d'entreprise dans quelques pays de l'OCDE en 1999



Notes : Le *taux de syndicalisation* est la proportion des salariés qui adhèrent à un syndicat (en %). La mesure de la *qualité du dialogue social* provient d'une enquête auprès de chefs d'entreprise. A la question : « Pensez-vous que les relations entre salariés et employeurs sont généralement coopératives ? », ils répondent par une note de 1 (pas du tout d'accord) à 7 (tout à fait d'accord). Les données présentées sont des moyennes par pays.

Source des données : OIT, OCDE, base de données *Global Competitiveness Report* 1999.

Source du graphique : P. Aghion, Y. Algan, P. Cahuc, 2008, « Can Policy Interact with Culture ? Minimum Wage and the Quality of Labor Relations », IZA discussion paper, n° 3680.

4. D'après la figure 3, le taux de syndicalisation en France (donnée entourée) est relativement faible.

   Vrai. Le graphique montre que le taux de syndicalisation en France s'élève à environ 10 %, ce qui est le niveau le plus faible parmi les pays de l'échantillon.

5. Figure 3 : Si la variable expliquée est le taux de syndicalisation, alors dans le cas de la France le résidu de la régression est négatif, c'est-à-dire que le modèle surestime le taux de syndicalisation en France.

   Vrai. Le point qui correspond à la France se situe en-dessous de la droite de régression. Autrement dit dans le cas de la France le taux de syndicalisation prédit par le modèle en fonction de l'autre variable, noté $\hat{y}_i$, est supérieur aux taux de syndicalisation observé, noté y_i. Alors le résidu, noté e_i est négatif : $e_i = y_i - \hat{y}_i < 0$

6. La covariance entre les deux variables de la figure 3 est positive et l'ajustement linéaire est presque parfait.

   Faux. La covariance est en effet positive car les deux variables sont corrélées positivement : plus le taux de syndicalisation est élevé, plus la qualité du dialogue social est bonne. Cela se traduit par la pente positive de la droite de régression. Par contre on ne peut pas parler d'un ajustement parfait parce que le coefficient de détermination est faible ($R^2 < 0{,}5$) ce qui se traduit graphiquement par l'existence de points qui sont éloignés de la droite de régression.

Exercice 3 : Le revenu nécessaire pour vivre

Chaque année, la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) du Ministère des affaires sociales et de la santé publie les résultats d'une enquête intitulée Baromètre d'opinions de la Drees. On pose notamment la question suivante aux individus interrogés : « Selon vous pour vivre, quel est le montant dont doit disposer AU MINIMUM un individu par mois ? » Le tableau 1 fournit des données concernant les réponses des individus à cette question.

Tableau 1 : Répartition des individus interrogés en fonction du montant déclaré nécessaire pour vivre (pour une personne, en euros par mois)

Année	Moins de 750 €	De 750 à moins de 900 €	De 900 à moins de 1100 €	De 1100 à moins de 1300 €	Plus de 1300 €	Ensemble
2000	19,4	23,5	21,4	25,5	10,2	100
2010	3,0	5,1	18,2	22,2	51,5	100

Source : BVA, Drees

Notre premier objectif est de calculer quelques indicateurs afin de résumer les données pour l'année 2000. **Nous posons les hypothèses suivantes** : le plus faible des montants donnés par les individus interrogés lors de l'enquête est zéro euro tandis que le plus élevé des montants donnés est 2050 €.

1. Calculez que la moyenne de cette série.
   On note X la variable « montant déclaré nécessaire pour vivre par personne et par mois (en euros) en 2000 ». $\overline{X} \approx 957{,}65$
2. Interprétez cette moyenne.
   Le montant moyen qui est déclaré nécessaire pour vivre par les Français interrogés en 2000 est 957,65 euros par personne et par mois.
3. Calculez le coefficient de variation du montant déclaré nécessaire pour vivre.
   $CV_X \approx 38{,}8\%$
4. Interprétez le coefficient de variation.
   Les individus interrogés donnent des montants minimums nécessaires pour vivre qui s'écartent en moyenne de 38,8 % par rapport à la moyenne.
5. Donnez la classe médiane. Montrez que la valeur médiane peut être approximée à 970.
   La classe médiane est la classe [900 ; 1100[car elle contient la valeur médiane. En effet 42,9 % des individus donnent des montants inférieurs à 900 € et 64,3 % des individus donnent des montants inférieurs à 1100 €. Le montant déclaré médian peut être approchée par interpolation linéaire : $\frac{1100-900}{0{,}643-0{,}429} = \frac{Me_X-900}{0{,}5-0{,}429} \Leftrightarrow Me_X = 900 + \frac{0{,}071}{0{,}214} * 200 \approx 966{,}4$
   En arrondissant ce résultat à la dizaine supérieure on obtient une médiane de 970 euros.
6. Interprétez la valeur médiane.
   La moitié des individus interrogés sur le montant nécessaire pour vivre par mois et par personne donne un chiffre supérieur à 970 euros, l'autre moitié donnant un chiffre inférieur.

Il s'agit dans un second temps de représenter graphiquement la distribution de la variable considérée en 2000, en conservant les hypothèses posées ci-dessus sur les valeurs extrêmes de cette distribution.

7. Calculez les fréquences corrigées. Pourquoi faut-il corriger les fréquences des différents intervalles de valeurs pour tracer l'histogramme ?
   Un histogramme est un graphique dont l'**aire** des rectangles est proportionnelle aux fréquences de chaque classe de valeurs. Il faut donc tenir compte des amplitudes variables selon les classes. Ainsi, pour tracer l'histogramme il nous faut calculer les densités de chaque classe de valeurs, notées d_i, définies comme le rapport entre la fréquence de la

classe et son amplitude : $d_i = \frac{f_i}{a_i}$; puis les fréquences corrigées que l'on obtient en multipliant les densités de chaque classe par la plus petite amplitude de classe : $f_i^{cor} = d_i * \min a_i$

8. Tracez l'histogramme. Utilisez le papier millimétré fourni en annexe, à détacher pour le joindre à votre copie.
9. *Question subsidiaire :* D'après votre graphique, quelle est la classe modale ? Expliquez.
   La classe modale est celle dont la densité ou la fréquence corrigée est la plus importante, c'est-à-dire la classe $[750\ ;\ 900[$. En tenant compte des amplitudes des différentes classes, un montant nécessaire pour vivre situé entre 750 et 900 euros est la réponse la plus fréquente.
10. *Question subsidiaire :* Si on considère que la distribution de la variable en 2000 est normale au sens statistique du terme, alors d'après la « règle empirique » quelle devrait être la proportion d'individus donnant un montant supérieur à 2070 € ? Supérieur à 1700 € ?
    Si la distribution est normale, on sait que la quasi-totalité des valeurs se situent dans l'intervalle $[\bar{X} - 3\sigma_X\ ;\ \bar{X} + 3\sigma_X]$. Or $\bar{X} + 3\sigma_X = 957{,}65 + 3 * 371{,}25 = 2071{,}4$ donc les réponses supérieures à 2070 euros doivent être exceptionnelles. De même la « règle empirique » associée à la loi normale nous apprend que seules 5 % des observations s'écartent de plus de deux écarts-types par rapport à la moyenne, c'est-à-dire que 95 % des observations se situent dans l'intervalle $[\bar{X} - 2\sigma_X\ ;\ \bar{X} + 2\sigma_X]$. On vérifie que $\bar{X} + 2\sigma_X \approx 1700$. Nous savons en outre qu'une distribution normale est parfaitement symétrique, ce qui nous permet de dire qu'environ 2,5 % des individus devraient donner un montant supérieur à 1700 euros.

Enfin, nous nous intéressons aux données pour 2010 et à la comparaison entre 2000 et 2010.

11. Comparez la moyenne de la série en en 2000 et en 2010. On conserve les mêmes hypothèses sur les valeurs extrêmes.
    On note Y la variable « montant déclaré nécessaire pour vivre par personne et par mois en 2010 (euros) ». $\bar{Y} \approx 1364{,}39$: le montant déclaré nécessaire pour vivre a augmenté d'environ 41 % en dix ans.
12. En 2010, le coefficient de variation s'élève à 103,9 %. En quoi ce résultat rend-il difficile l'interprétation de la valeur moyenne de 2010 ?
    Cela signifie que les montants donnés par les individus sont beaucoup plus dispersés en 2010 qu'en 2000. En effet, lorsque le coefficient de variation dépasse 100%, alors la moyenne a peu de sens.
13. Quelle serait le montant moyen en 2010 si on considère que le montant le plus élevé déclaré par les enquêtés est de 3 000 € ? En quoi l'hypothèse sur la valeur de la borne supérieure de la classe est-elle plus importante pour l'année 2010 que pour l'année 2000 ?
    On peut le calculer directement à partir du résultat de la question 11. On note $\bar{Y}'$ la moyenne obtenue avec le nouveau découpage, f la fréquence de la dernière classe, c le centre de la classe $[1300\ ;\ 2050]$ et c' le centre de la nouvelle classe $[1300\ ;\ 3000]$:
    $$\bar{Y}' = \bar{Y} - (f * c) + (f * c') \approx 1364{,}39 - 862{,}625 + 1107{,}25 \approx 1609{,}02$$
    On voit que la moyenne est fortement influencée par une augmentation de la borne supérieure de la dernière classe, ce qui tient à la fréquence élevée de cette classe (plus de la moitié des réponses). En 2000 cet effet est moins important car la fréquence de la dernière classe est plus faible. De ce point de vue le découpage par classe paraît mal adapté pour l'année 2010.
14. *Question subsidiaire :* À votre avis, est-ce une bonne idée de conserver le même découpage en intervalles de valeurs de 2000 à 2010 ? Expliquez.

C’est peu pertinent d’un point de vue économique car les prix ont augmenté en dix ans et donc le pouvoir d’achat d’une somme en euros courants a diminué. Par contre c’est pertinent d’un point de vue statistique : si on change les intervalles on ne peut plus directement comparer les réponses de 2000 et celles de 2010.