

École des sciences de la gestion – UQAM  
Département des sciences économiques  
Cours : ECO2273–50 Économétrie I  
Prof. : Philip Merrigan  
Trim. : automne 2021

**Travail pratique #1**  
(Une copie par étudiant, à remettre 22 octobre)

**Section 1**

**Statistiques descriptives (Problèmes 1, 2, 3, 4, 5, 6, 7, 8, répondez avec STATA, énoncez bien dans la réponse les commandes STATA retenues pour répondre aux problèmes)**

**Problème #1**

1001	1	7.2	Nord
1002	1	7.3	Nord
1003	2	10.1	Sud
2001	2	12	Est
2002	2	13	Sud
2003	1	6.8	Ouest
2021	2	7.9	Ouest
2022	1	9.2	Ouest
2033	1	11	Nord
2034	1	10.1	Est

La première colonne identifie une région, la seconde indique si la région est urbaine (1) ou rurale (2), la troisième le taux de chômage, la quatrième sa location dans le pays.

- a. Utilisez la commande input pour créer un fichier de données dans STATA, choisissez un nom pour chaque variable.
- b. Étiquetez chaque variable
- c. Étiquetez les valeurs de la variable de région dans la colonne 2
- d. Calculez avec une calculatrice, la moyenne et la variance du taux de chômage. Expliquez comment vous avez obtenu vos résultats. Calculez la médiane du taux de chômage sans STATA.

**Problème #2**

Le magazine *Fortune* établit une liste des plus importantes sociétés américaines en termes de chiffre d'affaires annuel. Le fichier **TP1-Grandes sociétés**, placé sur Moodle, présente le chiffre d'affaires annuel des 50 plus importantes entreprises américaines, exprimé en milliards de dollars (source : site Internet de *CNN Money*, 15 janvier 2010). Présentez toutes les distributions demandées dans les questions suivantes dans un seul tableau dont les lignes correspondent aux classes de chiffres d'affaires (0-49, 50-99, 100-149, 150-199, etc.) et les colonnes aux types de distribution demandés. Pour créer une variable représentant les différentes catégories de chiffre d'affaires, utilisez la

procédure STATA « egen » avec la fonction « cut ». Pour la fréquence cumulée utilisez la commande contract.

- a. Construire une distribution de fréquence des chiffres d'affaires annuels, de fréquence relative en pourcentage des chiffres d'affaires annuels et une distribution de fréquence cumulée relative des chiffres d'affaires annuels. Présentez le résultat de manière claire dans votre TP.
- b. Construire une distribution de fréquence cumulée des chiffres d'affaires annuels. Présentez le tableau dans le TP.
- c. Construire un histogramme des chiffres d'affaires annuels avec STATA en fraction d'abord puis en et en densité. Quelle est l'aire du plus grand rectangle de l'histogramme ? Expliquez votre raisonnement. Refaire avec un intervalle de 30 pour la base des rectangles. Donnez un titre à votre graphique.
- d. Quelle est la plus importante société américaine et quel est son chiffre d'affaires annuel ? Utilisez les commandes summarize et list avec l'option r() pour répondre à la question. En d'autres mots, vous devez programmer la réponse.

### Problème #3

Utilisez la banque de données sante.dta pour répondre aux questions sous l'onglet sante sur Moodle. Deux variables sont utilisées pour répondre aux questions. La première « smsa » indique si l'individu vit dans une zone urbaine et dans une ville centrale, dans une zone urbaine mais pas dans une ville centrale, pas dans une zone urbaine. La seconde est « highbp » qui indique si l'individu souffre d'hypertension.

- a. Renommez ces deux variables
- b. Étiquetez en français les valeurs de smsa (que vous avez renommé en français en a)
- c. Étiquetez en français les valeurs de highbp (que vous avez renommé en français en a)
- d. Faites un diagramme de barres de smsa (renommé en français)
- e. Faites un diagramme de barres de côte-à-côte de highbp selon les valeurs de smsa
- f. Refaites e, mais en barres empilées
- g. Produisez un tableau croisé des deux variables avec highbp en rangée et la région de résidence en colonne dites ce que chaque chiffre du tableau représente (choisissez l'ordre que vous désirez pour les variables).
- h. Produisez un tableau croisé pour les deux variables tel que les éléments de chaque rangée somment à 100 %. Expliquez ce que veulent dire les chiffres de la première rangée de ce tableau. Produisez le tableau mais en sommant les colonnes pour qu'elles arrivent à 100%.

### Problème #4

Lorsqu'on travaille avec un progiciel, il faut savoir lire la documentation sur les procédures. Cette question vous présente des procédures qui ne vous seront pas expliquées en classe.

La banque de données travail.dta (sur Moodle travail) est composée de travailleurs canadiens à qui l'on demande le nombre d'heures usuelles de travail (UTOTHRS). On a aussi dans la banque un identificateur pour l'individu (REC\_NUM). Si vous exécutez les 3 lignes du programme présenté au bas de la question, la première et la dernière procédure donne le même moyenne pour les heures usuelles de travail expliquez pourquoi ? Que fait la commande collapse ?

```
sum UTOTHRS
```

```
collapse(count) REC_NUM, by(UTOTHRS)
```

```
sum UTOTHRS [fweight = REC_NUM]
```

### Problème #5

**Répondez avec le fichier TP1-Billets Ligue nationale de foot.dta et STATA**

- Quel est le prix moyen du billet des 14 équipes.
- L'année précédente, le prix moyen du billet était de 72,20 dollars. Quelle a été l'augmentation moyenne, en pourcentage, du prix du billet sur un an ?
- Quel est le prix médian du billet. Calculez aussi la médiane du prix avec la règle vue classe pour un nombre d'observations pair.
- Quel sont les premier et troisième quartiles du prix du billet.
- Quelle est la variance et l'écart-type du prix du billet.
- Quelle est la valeur de la variable centrée réduite associée au prix du billet des Cowboys de Dallas ? Ce prix devrait-il être considéré comme une valeur aberrante ? Pourquoi ?
- Calculez avec la fonction `r()` l'étendue et l'étendue interquartile et le coefficient de variation du prix.
- Calculez limite inférieure et supérieure du prix des billets avec STATA en utilisant la fonction `r()` après `summarize`, montrez vos réponses avec la procédure `display`.

I Fabriquez la boîte-à-pattes du prix des billets

### Problème #6

Aux États-Unis, au début de l'année 2009, la crise économique a entraîné une augmentation sensible des pertes d'emplois et des saisies immobilières. Le taux de chômage national s'élevait à 6,5 % et le pourcentage de saisies immobilières à 6,12 % (*The Wall Street Journal*, 27 janvier 2009). Pour prévoir l'état du marché immobilier au cours de l'année à venir, les économistes ont étudié la relation entre le taux de chômage et le pourcentage de saisies immobilières. Les économistes pensaient que, si le taux de chômage continuait à augmenter, il y aurait également une augmentation des saisies immobilières. Les données contenues dans le fichier **TP1-Logement**, placé sur Moodle, présentent le taux de chômage et le pourcentage des saisies immobilières sur 27 marchés immobiliers américains.

- Calculer les écarts-type du taux de chômage et du pourcentage des saisies immobilières de cet échantillon de marchés immobiliers.
- Calculer la covariance et le coefficient de corrélation entre le taux de chômage et le pourcentage des saisies immobilières de cet échantillon de marchés immobiliers. Quelles sont les formules utilisées par STATA pour calculer ces statistiques.
- Y a-t-il une corrélation positive entre le taux de chômage et le pourcentage des saisies immobilières ? Dans l'affirmative, comment interprétez-vous cette corrélation ?
- Dessiner un nuage de points et une droite de tendance de la relation entre le taux de chômage et le pourcentage de saisies immobilières de l'échantillon.

### Problème #7

Les indices Dow Jones Industriel (DJIA) et Standard & Poor's 500 (S&P500) sont des indicateurs de l'évolution des prix des actions sur le marché boursier. L'indice DJIA est basé sur les variations de prix des 30 plus grandes sociétés américaines ; l'indice S&P500 est composé de 500 titres. Certains prétendent que le S&P500 est un meilleur indicateur de la performance des marchés boursiers dans la mesure où il suit l'évolution d'un plus grand nombre de titres. Les prix de clôture des indices DJIA et S&P500 durant 15 semaines, du vendredi 6 janvier 2012 au vendredi 13 avril 2012, sont présentés dans le fichier **TP1-DJIAS&P500**, placé sur Moodle, (source : site Internet de *Barron's*, 17 avril 2012).

Supposez que vous déteniez un portefeuille de 200 000 dollars composé à 50 % de titres dont le rendement suit exactement l'indice DJIA et à 50 % de titres dont le rendement suit exactement l'indice S&P500. On vous demande de calculer le **rendement hebdomadaire moyen** de chacune des composantes de votre portefeuille au cours de la période de **14 semaines** débutant le 9 janvier à l'ouverture de la Bourse de New York et se terminant le vendredi 13 avril 2012. Notez que le cours de chacun des indices à l'ouverture des marchés boursiers du 9 janvier correspond au cours de fermeture des marchés du vendredi 6 janvier.

- Calculer le rendement hebdomadaire moyen de chacune des composantes de votre portefeuille en utilisant toutes les informations disponibles (c'est-à-dire les valeurs des indices boursiers aux 15 dates présentées dans le fichier).
- Calculer le rendement hebdomadaire moyen de chacune des composantes de votre portefeuille en n'utilisant que les valeurs des indices boursiers du 6 janvier et du 13 avril 2012.

Conseil : comme les taux de rendement hebdomadaire obtenus sont relativement faibles, conservez suffisamment de décimales dans tous vos calculs pour que vos approximations ne soient pas une source d'erreur importante.

### Problème #8

Le fichier **TP1-Universités.dta**, placé sur Moodle, présente les données relatives à un échantillon de 103 établissements d'enseignement privés (collèges et universités) américains. Les données comprennent le nom du collège ou de l'université, l'année de la création de l'établissement, les droits de scolarité (hors pension) au cours des années les plus récentes ainsi que le pourcentage des étudiants de l'établissement qui ont

obtenu leur diplôme de baccalauréat en six ans ou moins (source : *The World Almanac*, 2012).

### Section A

- A1. Construire une tabulation croisée avec l'année de création de l'établissement en ligne et les droits de scolarité en colonne. Utiliser des classes de 50 ans commençant en 1600 et se terminant en 2000 pour l'année de création. Pour les droits de scolarité, utiliser des classes de 5 000 dollars commençant à 1 dollar et se terminant à 45 000 dollars.
- A2. Calculer les pourcentages en ligne pour la tabulation croisée développée à la sous-question A1.
- A3. Quelle relation, s'il en existe une, remarquez-vous entre l'année de la création de l'établissement et le niveau de ses droits de scolarité ?

### Section B

- B1. Construire une tabulation croisée avec l'année de création de l'établissement en ligne et le pourcentage de diplômés en colonne. Utiliser des classes de 50 ans commençant en l'an 1600 et se terminant en l'an 2000 pour l'année de création. Pour le pourcentage de diplômés, utiliser des classes de 5 % commençant à 35 % et finissant à 100 %.
- B2. Calculer les pourcentages en ligne pour la tabulation croisée développée à la sous-question B1.
- B3. Commentez la relation, s'il en existe une, entre ces variables.

### Section C

- C1. À l'aide d'un nuage de points et d'une droite de tendance, représenter la relation entre l'année de création de l'établissement et le niveau des droits de scolarité.
- C2. Commenter la relation entre les deux variables.

### Section D

- D1. À l'aide d'un nuage de points et d'une droite de tendance, représenter la relation entre le niveau des droits de scolarité et le pourcentage de diplômés.
- D2. Commenter la relation entre les deux variables.

## Probabilités

### Problème #1

La tabulation croisée suivante présente les revenus des ménages américains selon le niveau d'études de leur chef (*Statistical Abstract of the United States*, 2008).

Niveau d'études du chef	Revenus des ménages (en milliers de dollars)					Total
	< 25	25-49,9	50-74,9	75-99,9	100 +	
High School sans diplôme	4 207	3 459	1 389	539	367	9 961
Diplômé du High School	4 917	6 850	5 027	2 637	2 668	22 099
Université sans diplôme	2 807	5 258	4 678	3 250	4 074	20 067
Baccalauréat	885	2 094	2 848	2 581	5 379	13 787
Maîtrise ou Doctorat	290	829	1 274	1 241	4 188	7 822
Total	13 106	18 490	15 216	10 248	16 676	73 736

- a. Construire le tableau des probabilités jointes et des probabilités marginales.
- b. Quelle est la probabilité qu'un chef de ménage ne possède pas un diplôme de high school ?
- c. Quelle est la probabilité qu'un chef de ménage possède un diplôme égal ou supérieur au baccalauréat ?
- d. Quelle est la probabilité que le revenu d'un ménage dirigé par une personne diplômée d'un baccalauréat soit de 100 000 dollars ou plus ?
- e. Quelle est la probabilité que le revenu d'un ménage soit inférieur à 25 000 dollars ?
- f. Quelle est la probabilité que le revenu d'un ménage dirigé par une personne diplômée d'un baccalauréat soit inférieur à 25 000 dollars ?
- g. Les revenus des ménages américains sont-ils statistiquement indépendants du niveau d'études de leur chef ?

### Problème #2

Une étude portant sur 31 000 admissions dans les hôpitaux de l'État de New York a permis d'évaluer à 4 % le nombre d'admissions suivies d'infections dues aux traitements. Un septième de ces infections avaient entraîné le décès du malade et un quart étaient le résultat de négligence. Dans un cas sur 7,5 impliquant de la négligence, une plainte pour faute professionnelle (*malpractice*) avait été déposée et des dédommagements financiers avaient été obtenus une fois sur deux.

Pour y voir plus clair, définissez les événements suivants : (I) : le patient subit une infection à la suite d'un traitement ; (D) : le patient décède à la suite d'une infection ; (N) : l'infection résulte d'une négligence ; (M) : le patient poursuit l'hôpital pour faute professionnelle ; (S) : le patient obtient un dédommagement financier.

- a. À l'aide des informations dont vous disposez, évaluez les probabilités suivantes : (1)  $P(I)$  ; (2)  $P(N|I)$  ; (3)  $P(D|I)$  ; (4)  $P(M|N)$  ; (5)  $P(S|M)$ .
- b. Quelle est la probabilité qu'une personne admise à l'hôpital souffre d'une infection causée par de la négligence ?
- c. Quelle est la probabilité qu'une personne admise à l'hôpital meure à la suite d'une infection causée par de la négligence ?
- d. Dans les cas d'infections causées par de la négligence, quelle est la probabilité qu'une plainte pour faute professionnelle résulte en un dédommagement financier ?

### Problème #3

Une entreprise s'interroge sur la rentabilité de la mise en marché d'un nouveau produit. À première vue, elle considère que les probabilités que les ventes du produit soient faibles (VF), modérées (VM) et élevées (VE) sont respectivement 0,40, 0,40 et 0,20. Elle procède d'abord à une mise en marché du produit dans une région test. Le niveau des ventes réalisées à cet endroit s'avère modéré. Par ailleurs, les données historiques relatives au niveau des ventes dans cette région indiquent que : (1) lorsque les ventes sont faibles au niveau national, elles sont modérées dans la région test 30 % du temps ; (2) lorsque les ventes sont modérées au niveau national, elles sont aussi modérées dans la région test 70 % du temps et, (3) lorsque les ventes sont élevées au niveau national, elles sont modérées dans la région test 10 % du temps. Comment l'information qui précède permettra-t-elle à l'entreprise de réviser ses prévisions relatives à la rentabilité du nouveau produit ? Quelles seront ces prévisions après révision ?

#### Problème #4

Un responsable de programme universitaire a constaté que 20 % de ses étudiants quittaient le programme sans avoir réussi le cours d'introduction aux statistiques. Supposons que 20 étudiants se soient inscrits à ce cours ce trimestre.

- Quelle est la probabilité qu'au plus deux étudiants quittent le programme ?
- Quelle est la probabilité qu'exactement quatre étudiants quittent le programme ?
- Quelle est la probabilité que plus de trois étudiants quittent le programme ?
- Quelle est l'espérance mathématique du nombre de départs ?

#### Problème #5

Santé ! Le minibar d'une chambre d'hôtel révèle généralement si l'hôtel est haut de gamme ou non. Par ailleurs, les travaux du centre *PKF Hospitality* ont montré que les consommations des minibars fournissaient un revenu annuel moyen de 368 dollars par chambre (*USA Today*, 9 février 2012). Considérez un hôtel haut de gamme de San Antonio, Texas, doté au total de 330 chambres, chacune disposant d'un minibar. Supposez que le revenu mensuel total du service minibar de l'hôtel suive une loi normale avec un écart-type de 2 200 dollars.

- Sachant que le revenu annuel moyen par minibar est de 368 dollars, quel est le revenu mensuel total moyen du service minibar de l'hôtel ?
- Quelle est la probabilité que le service minibar de l'hôtel génère un revenu mensuel supérieur à 12 000 dollars ?
- Quelle est la probabilité que le service minibar de l'hôtel génère un revenu mensuel inférieur à 7 500 dollars ?
- L'hôtel étudie la possibilité de proposer des boissons plus haut de gamme pour rendre le minibar plus attrayant. La direction de l'hôtel estime que les nouvelles offres du minibar pourraient accroître le revenu annuel moyen jusqu'à 420 dollars par minibar. Supposez que le revenu mensuel total du nouveau service de minibar de l'hôtel suive une loi normale avec un écart-type de 2 500 dollars. Répondez de nouveau aux sous-questions (b) et (c) pour le service amélioré de minibar. Soutiendriez-vous la stratégie de montée en gamme du service de minibar de l'hôtel ? Pourquoi ?

#### Problème #6

L'Association de contrôle et d'audit des systèmes d'information (ACASI) a mené une enquête auprès d'employés de bureau pour évaluer l'intensité de l'usage qu'ils feraient de leur ordinateur professionnel pour effectuer leurs achats de Noël. Supposez que le nombre d'heures passées par un employé à effectuer des achats de Noël avec son ordinateur professionnel suive une loi exponentielle.

- Le rapport de l'ACASI a conclu qu'il y a une probabilité de 0,53 qu'un employé n'utilise pas plus de 5 heures son ordinateur professionnel pour effectuer des achats de Noël. Est-ce que le temps moyen passé à effectuer des achats de Noël avec un ordinateur professionnel est plus près de 5,8, 6,2, 6,6 ou 7 heures ?
- À l'aide du temps moyen calculé en (a), évaluez la probabilité qu'un employé passe plus de 10 heures à effectuer des achats de Noël avec son ordinateur professionnel.

- c. Quelle est la probabilité qu'un employé utilise son ordinateur professionnel entre 4 et 8 heures pour effectuer des achats de Noël ?

### Problème #7

La compagnie PortaCom a récemment mis au point une imprimante portable de haute qualité. Les deux composantes principales du coût de fabrication de l'appareil sont le coût de la main-d'œuvre et le coût des pièces. Durant la période d'essai, l'entreprise a effectué des tests complets avec la nouvelle imprimante. De plus, les ingénieurs ont obtenu la distribution de probabilité bivariable suivante des coûts de fabrication de l'appareil. Le coût des pièces (en dollars par imprimante) est représenté par la variable  $x$  et le coût de la main-d'œuvre (en dollars par imprimante) par la variable  $y$ . La direction de l'entreprise compte utiliser cette distribution de probabilité pour évaluer les coûts de fabrication de ses imprimantes.

Pièces ( $x$ )	Main-d'œuvre ( $y$ )		
	43	45	48
85	0,05	0,20	0,20
95	0,25	0,20	0,10

- Évaluer la distribution de probabilité marginale du coût de la main-d'œuvre et calculer son espérance, sa variance et son écart-type.
- Évaluer la distribution de probabilité marginale du coût des pièces et calculer son espérance, sa variance et son écart-type.
- Sachant que le coût total de fabrication d'une imprimante est la somme du coût de la main-d'œuvre et du coût des pièces, présenter la distribution de probabilité du coût total de fabrication d'une imprimante.
- Calculer l'espérance, la variance et l'écart-type du coût total de fabrication d'une imprimante.
- Le coût de la main-d'œuvre et le coût des pièces sont-ils des variables indépendantes ? Répondez à cette question en comparant les probabilités conditionnelle et marginale appropriées.
- Le coût de la main-d'œuvre et le coût des pièces sont-ils des variables indépendantes ? Répondez à cette question en calculant le coefficient de corrélation entre ces deux variables.
- Si le coût de la main-d'œuvre et le coût des pièces ne sont pas des variables indépendantes, quelle relation semble-t-il exister entre elles ? Pourquoi ?

### Problème #8

Supposez que les revenus annuels de travail de chacun des conjoints, tous deux présents sur le marché du travail, dans une population de ménages puissent être répartis en trois classes : revenus faibles (20 000 \$), moyens (60 000 \$) et élevés (100 000 \$). Supposez également que le tableau A soit une représentation fidèle des probabilités de choisir, de façon aléatoire, chacun des neuf types de ménages présents dans la population. Les variables  $R_H$  et  $R_F$  représentent les revenus annuels de travail des hommes et des femmes ; elles sont exprimés en milliers de dollars.

A		$R_H$		
		20	60	100
	20	0,10	0,10	0,10



R <sub>F</sub>	60	0,10	0,20	0,10
	100	0,10	0,10	0,10

- Calculer la moyenne, la variance et l'écart-type des revenus de travail de chacun des membres des ménages (considérés séparément).
- Calculer la moyenne, la variance et l'écart-type des revenus de travail des ménages.
- Supposez maintenant que le tableau B présente une répartition des revenus de travail des ménages légèrement différente au sein de la population. Les revenus de travail des ménages seraient-ils alors répartis de façon plus égalitaire ou moins égalitaire que dans le tableau A ? Justifiez votre réponse par un calcul de variance.

B		R <sub>H</sub>		
		20	60	100
R <sub>F</sub>	20	0,05	0,05	0,20
	60	0,05	0,30	0,05
	100	0,20	0,05	0,05

- Supposez enfin que le tableau C présente une troisième répartition des revenus de travail des ménages au sein de la population. Les revenus de travail des ménages seraient-ils alors répartis de façon plus égalitaire ou moins égalitaire que dans le tableau A ? Justifiez votre réponse de façon rigoureuse par un calcul de variance.

C		R <sub>H</sub>		
		20	60	100
R <sub>F</sub>	20	0,20	0,05	0,05
	60	0,05	0,30	0,05
	100	0,05	0,05	0,20

Source : Anderson et coll., *Statistiques pour l'économie et la gestion*, De Boeck, 2015.