L1-MATH - STATISTIQUES DESCRIPTIVES

FEUILLE DE TRAVAUX DIRIGÉS N° 2

séries bi-variées

Enseignants : H. El-Otmany & V. Darrigrand **A.U.** : 2014-2015

1. Variables non-numériques

Exercice n°1 L'étude palynologique d'un échantillon du Mésozoïque d'Aquitaine a permis de construire la table de contingence suivante :

Richesse \ Couleur	Gris foncé	Grisâtre	Rougeâtre
Non fossilifère	26	19	18
Fossilifère	34	9	4

1. Préciser la population étudiée et la nature des deux variables.
2. Calculer les effectifs marginaux des deux distributions. Calculer l'effectif total.
3. Déterminer la distribution de la richesse des roches conditionnellement au fait que la couleur des roches est gris foncé.
4. Déterminer la distribution de la couleur des roches conditionnellement au fait qu'elles ne sont pas fossilifères.
5. Déterminer le mode de la distribution bivariée.
6. Peut-on considérer que les deux variables étudiées sont indépendantes ?

Exercice n°2 Une distribution de fréquences $\{(x_i, y_i), f_{ij}\ ; i = 1, 2\ ; j = 1, \ldots, 4\}$ de deux caractères non numériques observés sur une population est représentée par le tableau ci-dessous. On suppose que les modalités ont été codées : $E = \{20, 40\}$ (pour la variable x) et $F = \{10, 20, 30, 40\}$ (pour la variable y).

x \ y	10	20	30	40
20	0.04	0.08	0.08	0.05
40	0.12	0.24	0.24	0.15

Les deux variables sont-elles indépendantes ? Était-il possible de le voir rapidement sur le tableau ?

Exercice n°3 Le tableau ci-dessous indique pour chaque couple de modalités de deux caractères non numériques, la fréquence d'apparition de ce couple dans une population donnée. On suppose que les modalités ont été codées : $E = \{5, 15, 25\}$ (pour la variable x) et $F = \{10, 20, 30\}$ (pour la variable y). Certaines de ces données ont été remplacées par les paramètres a, b et c.

x \ y	10	20	30
5	0.09	a	0.06
15	0.15	0.25	b
25	c	0.10	0.04

1. Calculer le coefficient de Pearson Φ^2 en fonction de a, b et c.
2. Déterminer les réels a, b et c sachant que les deux variables x et y sont indépendantes.
3. Donner la distribution conditionnelle de y sachant $x = 5$.

2. Variables numériques

Exercice n°4 Une population de n ménages où les deux conjoints occupent un emploi est étudiée. Les variables x et y désignent respectivement les revenus du mari et de la femme. Une étude a montré que $\sigma_x^2 = 16$, $\sigma_y^2 = 10$ et $\sigma_{xy} = 8$. Calculer la variance du revenu total $s = x + y$ et du revenu disponible après impôt $w = 0.8x + 0.7y$.

Exercice n°5 On considère une automobile roulant sur une route mouillée à une certaine vitesse v. Cette automobile freine brusquement pour simuler un freinage d'urgence. On note d la distance de freinage qui est fonction de la vitesse v. Les données recueillies sont les suivantes :

Vitesse v (Km/h)	50	60	70	80	90	100	110
Distance d (m)	10	21	34	49	66	85	106

1. Quelles sont les natures des deux variables étudiées ?
2. Tracer le nuage de points. Le nuage de points vous semble-t-il aligné le long d'une droite ? Quel est le signe du coefficient de corrélation linéaire (sans effectuer de calculs) ?
3. On suppose qu'il existe une relation quadratique entre les deux variables, i.e. une relation de la forme $d = av^2 + b$.
   (a) Quelles sont les unités de a et b.
   (b) Déterminer a et b par la méthode des moindres carrés.
   (c) Représenter la courbe obtenue sur le graphique précédent.
   (d) Quelle est la distance estimée de freinage si la vitesse est de 95 Km/h ? 130 Km/h ? 150 Km/h ?
   (e) Que deviennent a et b si on exprime la distance en kilomètres et non pas en mètres ?
   (f) Que vaut le coefficient de corrélation linéaire ? le coefficient de détermination ?
4. Trois autres expériences supplémentaires ont été menées :

Vitesse v (Km/h)	110	120	130
Distance d (m)	129	154	181

   (a) Déterminer les nouvelles valeurs de a et de b par la méthode des moindres carrés, en limitant le nombre de calculs. On expliquera la démarche adoptée.
   (b) Que vaut le coefficient de détermination ? Comparer avec celui précédemment obtenu et commenter.
   (c) Quelle est la distance estimée de freinage si la vitesse est de 95 Km/h ? 130 Km/h ? 150 Km/h ? Comparer avec les résultats obtenus précédemment et commenter.

Exercice n°6 On donne pour les six premiers mois de l'année 1982 les nombres d'offres d'emploi (concernant les emplois durables et à plein temps) et le nombre des demandes d'emploi (déposées par des personnes sans emploi, immédiatement disponibles et à la recherche d'un emploi durable et à plein temps). Les données sont exprimées en milliers d'individus.

Offres (x)	61	66.7	75.8	78.6	82.8	87.2
Demandes (y)	2034	2003.8	1964.5	1928.2	1885.3	1867.1

1. Représenter le nuage de points. Le nuage de points vous semble-t-il aligné le long d'une droite ?
2. Trouver la droite de régression des demandes d'emploi en fonction des offres d'emploi et la tracer sur le graphique précédent.
3. Calculer le coefficient de corrélation entre x et y. Commenter.

Exercice n°7 On donne pour les années 1975 à 1983, le prix du ticket de métro à Paris acheté à l'unité en août (x) et le prix moyen annuel du kilogramme de bananes dans la région parisienne (y).

Prix ticket (x)	2.2.	2.5	2.7	3	3.6	4.5	5	5.5	6
Prix Kg de bananes (y)	3.72	3.95	4.27	4.52	5.01	5.41	6.17	7.03	8.42

Répondre aux mêmes questions qu'à l'exercice précédent. Commenter.

Exercice n°8 Le tableau ci-dessous donne les valeurs expérimentales du volume V (en cm^3) et de la pression P (en Kg par cm^3) d'un gaz. D'après les lois de la thermodynamique de Laplace pour un gaz parfait, on a la relation $PV^\gamma = C$ où γ et C sont des constantes.

Volume (v)	620	890	1013	1186	1454	1944	2313	3179
Pression (p)	6.7	4.3	3.48	2.644	1.997	1.35	1.1	0.71

1. Transformer le modèle pour obtenir un modèle linéaire.
2. Déterminer la droite de régression linéaire pour le modèle transformé. En déduire une estimation $\hat{\gamma}$ et $\hat{C}$ des constantes γ et C.
3. Estimer P lorsque le volume $v = 1000$ cm^3.
4. Soit $\hat{u}_i = p_i - \hat{p}_i$ les écarts entre les observations et les valeurs ajustées. Montrer que u_i est approximativement égale à $\hat{p}_i \hat{\varepsilon}_i$ où $\hat{\varepsilon}_i$ désigne l'erreur pour le modèle transformé. Vérifier numériquement sur les données.

Exercice n°9 Le taux d'équipement des ménages en matériel informatique est une variable y_t où t représente l'année de l'observation. On fait l'hypothèse d'un modèle logistique :

$$y_t = \frac{1}{1 + ae^{-bt}} ,$$

avec a une constante positive.

1. Identifier les individus de cette étude.
2. Par un changement de variable approprié, montrer que le modèle logistique peut être transformé en un modèle linéaire que l'on précisera.
3. Appliquer la méthode des moindres carrés sur le modèle transformé en utilisant les données ci-dessous :

Années	1988	1989	1990	1991	1992	1993	1994
t	1	2	3	4	5	6	7
y	0.45	0.57	0.69	0.78	0.86	0.91	0.93

4. Tracer la droite de régression du modèle transformé ainsi que le nuage de points correspondant.
5. En déduire une estimation des paramètres a et b. Sur un autre graphique, tracer la courbe estimée et le nuage de points des données initiales. Le modèle logistique vous semble-t-il bien spécifié ?
6. Prévoir le taux d'équipement en 1996. En quelle année le taux d'équipement sera-t-il de 99% ?

Exercice n°10 Une étude de marché a permis de relever le volume (y) des ventes (en milliers d'euros) d'un produit en fonction de son prix (x) de vente (en euros) :

x	95	130	146	210	250	280	330	350
y	104	58	37	22	12	12	9	7

1. Donner le signe du coefficient de corrélation linéaire ρ_{xy} sans effectuer de calculs. En déduire le signe du coefficient de corrélation linéaire $\rho_{\ln(x)\ln(y)}$, toujours sans effectuer de calculs. Calculer le coefficient de corrélation de Spearman.
2. On fait l'hypothèse d'un modèle de la forme $y = kx^{-m}$ ($k > 0$). Le coefficient m s'appelle, en économie, le coefficient d'élasticité. Estimer ce coefficient en se ramenant à un modèle linéaire. Prévoir le volume des ventes si le prix de vente est fixé à 360 euros.
3. Quel aurait été l'effet sur l'estimation des coefficients k et m pour la même étude mais avec des prix exprimés en francs ?

Exercice n°11 Dans douze familles, on a relevé le quotient intellectuel du père (x) et celui du fils (y) :

x	123	144	105	110	98	138	131	90	119	109	125	100
y	102	138	126	133	95	146	115	100	142	105	130	120

1. Représenter le nuage de points.
2. Existe-t-il une relation linéaire entre les deux variables ? Si oui, tracer la droite de régression.
3. Existe-t-il une relation non linéaire entre les deux variables ?

Exercice n°12 Un professeur émet l'hypothèse d'une liaison entre les notes à un test (x) et le temps (y) en minutes consacré au travail effectué quotidiennement à la maison. Sur huit élèves, les résultats sont les suivants :

x	8	12	4	6	15	9	6	18
y	14	15	8	7	20	11	10	17

1. Existe-t-il une relation linéaire entre les deux variables ?
2. Existe-t-il une relation non linéaire entre les deux variables ?

3. Variables numériques et variables non-numériques

Exercice n°13 On a mesuré (en mm) la longueur des œufs de coucous trouvés dans les nids de deux espèces d'oiseaux. On a obtenu les résultats suivants :

— nids de petite taille (roitelets) :

19.8	22.1	21.5	20.9	22.0	21.0	22.3	21.0
20.3	22.0	22.0	20.9	20.8	21.2	21.0	21.3

— nids de grande taille (fauvettes) :

22.0	23.9	20.9	23.8	25.0	24.0	23.8	21.7
22.8	23.1	23.5	23.0	23.0	23.1	23.2	

1. Quelle est la population étudiée ? Quelles sont les deux variables étudiées ? Préciser leur nature.
2. Représenter les deux distributions par deux histogrammes sur le même graphique.
3. Représenter les deux distributions à l'aide de boîtes à moustaches.
4. Calculer la variance intra-classe et la variance inter-classe.
5. Commenter.