Univsersité de Perpignan - IUT de Carcassonne
Thierry SPINOSA - Vivien ROSSI
Année 2006/2007

IUT STID, 2ème année : **Estimation et tests**

Feuille 5 : Tests du χ^2 et de corrélation

Exercice 1

100 candidats se sont présentés à deux concours A et B. 30 candidats ont été reçus au concours A, 60 ont été reçus au concours B et 23 candidats ont échoués aux deux concours. Nous nous interrogeons sur la liaison entre les réussites à ces deux concours.

1. Posez les hypothèses
2. Représenter la distribution dans un tableau
3. Effectuez un test du χ^2 et concluez au seuil d'erreur $\alpha = 5\%$

Exercice 2

Sur les 165 élèves de seconde d'un lycée qui ont subi deux épreuves d'évaluation (une en math et une en français), la répartition des notes est la suivante : 63 ont eu la moyenne aux deux épreuves, 32 ont eu la moyenne en français et pas en math, 12 ont eu la moyenne en math et pas en français et 58 n'ont pas eu la moyenne aux deux épreuves.

Peut-on dire que les deux épreuves ont un niveau de difficulté différent ? (vous testerez l'indépendance au seuil de 1%).

Exercice 3

Les données sont les offres de retransmission sportives par chaîne de télévision (TF1, A2, FR3, LA5) exprimées en millième pour différentes disciplines lors de l'année 1990 (Foot, Rugby, Tennis, Sport d'hiver, Cyclisme, Golf, Sport automobile, Athlétisme, Autres sports). On souhaite construire le test d'indépendance entre les variables «chaînes» et «sport».

Sport	TF1	A2	FR3	LA5
Foot	595	139	104	58
Rugby	0	132	13	16
Tennis	0	294	547	547
Hiver	37	44	69	0
Vélo	0	171	29	153
Golf	0	14	35	0
Auto.	166	2	15	208
Athlé.	1	49	18	0
Autres	201	155	170	18

1. Le tableau ci-dessus est un «tableau de profils». Lequel ?
2. Le tableau suivant est le tableau construit sous l'hypothèse d'indépendance. Compléter la première colonne.

Sport	TF1	A2	FR3	LA5
Foot	224.00			
Rugby		40.25		
Tennis			347.00	
Hiver				
Vélo				
Golf				12.25
Auto.	97.5			
Athlé.			17	
Autres				

Quelle particularité possède-t-il compte tenu de la question 1 ?

3. Le tableau suivant est le tableau des contributions de chaque croisement à la distance de χ^2 :

Sport	TF1	A2	FR3	LA5
Foot	?	32.2545	64.2857	123.0179
Rugby	40.2500	209.1444	18.4488	14.6102
Tennis	347.0000	8.0951	115.2738	115.2738
Hiver	0.0067	1.1267	26.4600	37.5000
Vélo	88.2500	77.5928	39.7797	47.5078
Golf	12.2500	0.2500	42.2500	12.25
Auto.	47.6528	93.7909	70.0518	124.3485
Athlé.	15.0588	60.2353	0.0588	17.0000
Autres	31.0662	2.6544	8.5000	102.3824

   (a) Retrouver la valeur manquante ?

   (b) Quel est le degré de liberté de la distance du χ^2 associée à ce tableau ?

   (c) Peut-on considérer que les deux variables «chaînes» et «sport» sont indépendantes ? (niveau $\alpha = 5\%$)

Exercice 4

On dénombre quotidiennement, pendant 100 jours, les accidents qui se sont produits dans une même ville. Les résultats expérimentaux sont résumés dans le tableau suivant :

Nb quotidien d'accidents	0	1	2	3 ou +	Total
Nb observés de jours	35	40	17	8	100

Vérifier (au niveau 5%) si le nombre quotidien d'accidents d'automobile obéit à une loi de Poisson de paramètre 1.

Rappel : Soit une variable aléatoire X de loi de Poisson de paramètre p, pour tout entier n on a $\mathbb{P}(X = n) = \frac{\exp(-p)p^n}{n!}$.

Exercice 5 : Loi de Mendel

Gregor Mendel (1822-1884) choisit le poids comestible comme objet de ses expériences et s'intéressa à deux caractères :
- la couleur : jaune (J) ou vert (V)
- la forme : ronde (Ro) ou ridée (Ri)

En croisant des lignées pures de pois ayant des graines jaunes et rondes avec des lignées pures des pois ayant des graines vertes et ridées, il constata que les pois hybrides ainsi créés avaient des graines jaunes et rondes. Il conclut que la couleur jaune devait être dominante sur la couleur verte et la forme ronde dominante sur la forme ridée. Pour certaines raisons biologiques, il conclut donc que si l'on croise les pois hybrides entre eux, on doit avoir les probabilités suivantes pour leurs graines : $\mathbb{P}(J \,\&\, Ro) = 9/16$,$P(J \,\&\, Ri) = 3/16$, $\mathbb{P}(V \,\&\, Ro) = 3/16$ et $\mathbb{P}(V \,\&\, Ri) = 1/16$. Ce résultat s'appelle la théorie de Mendel. Après avoir croisé les pois hybrides entre eux, il obtient 556 graines réparties comme suit :

Graines	Jaunes et rondes	Jaunes et ridées	Vertes et rondes	Vertes et ridées
Effectifs obs	315	101	108	32

Vérifiez (au niveau 10%) la théorie de Mendel.

Exercice 6

Un lac est peuplé par 4 espèces de salmonidés. A l'équilibre écologique, la répartition par espèce est la suivante :

Salmo trutta fario	Salmo trutta lacustris	Salmo trutta macrostigma	Salvelinus fontinalis
44%	44%	8%	4%

En fin de saison, on comptabilise le nombre de poissons de chaque espèce pris par les pêcheurs. On obtient le tableau suivant :

Salmo trutta fario	Salmo trutta lacustris	Salmo trutta macrostigma	Salvelinus fontinalis
357	312	50	40

Peut-on considérer que les prises de pêche sont en accord avec l'équilibre écologique ?

Exercice 7

On a noté dans cinq hôpitaux, parmi les personnes atteintes d'une certaine maladie (notées M), les nombres de guérisons dues à un traitement particulier. Les résultats sont les suivants :

Hôpital	1	2	3	4	5
Nombre de malades (M)	22	18	15	33	10
Nombre de guérisons	11	6	7	24	5

Peut-on admettre que le taux de succès est le même dans tous les groupes ? (niveau 5%)

Exercice 8

On veut tester l'efficacité d'un sérum contre une maladie. On considère 2 échantillons indépendants de 100 personnes chacun, toutes atteintes de cette maladie. Après un certain délai, on a noté 75 guérisons dans l'échantillon des traités et 65 dans celui des non traités.

1. Présenter les données dans une table de contingence.
2. Peut-on conclure à l'efficacité du sérum ? (niveau 5 %)
   – on formulera les hypothèses testées,
   – on rappellera les conditions d'application du test,
   – on interprètera les résultats obtenus.

Exercice 9

Le Titanic a emporté à son bord 325 passagers en première classe, 285 en seconde classe, 706 en troisième classe et 885 membres d'équipage. Parmi les survivants, on comptait 203 passagers de première classe, 118 de seconde classe, 178 de troisième classe et 212 membres d'équipage. Y-a-t-il un lien entre la classe et le fait d'avoir survécu ?

Exercice 10

On considère les résultats d'un groupe de 200 étudiants à un examen (notes de 0 à 20) :

Note	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Nb étudiants	1	5	10	15	20	27	33	30	20	15	12	7	4	1

Peut-on admettre, au niveau 10% que l'on a affaire à une loi normale .

Exercice 11

La répartition des durées (en heures) de 670 vols Paris-Alger en Caravelle est donnée dans le tableau suivant :

Durée	1.9		1.95		2		2.05		2.1		2.15		2.2		2.25		2.3		2.35		2.4		2.45		2.5		2.55
Nombre		19		19		39		48		87		94		104		92		57		44		28		26		13	

Est-ce que la durée du vol suit une loi normale ?

Exercice 12

Lors d'un examen fonctionnel respiratoire, le volume expiratoire moyen par seconde (VEMS) est mesuré sur 10 sujets avec deux appareils différents A_1 et A_2. Les valeurs obtenues (en litres) sont notées respectivement x et y et figurent dans le tableau suivant :

Sujets	1	2	3	4	5	6	7	8	9	10
$A_1 : x$	3.5	3.8	4.1	3	3.2	3	2.8	3.5	3.4	4
$A_2 : y$	3.3	3.5	4	3.1	3	3.5	3	3.5	3.2	3.6

1. Existe-il une corrélation linéaire significative entre les deux variables x et y ?
2. Le médecin ayant obtenu les mesures ci-dessus désire savoir si les deux appareils donnent des résultats différents. Que lui répondre ? (justifier votre réponse)

Tous les tests seront réalisés au niveau 5%.

Exercice 13

Au cours d'une étude quantitative consacrée à une race donnée de vaches laitières, on a observé diverses caractéristiques des animaux et de leur production. Nous considérons ici quelques valeurs relatives à la longueur du corps (variable X, mesurée en centimètre) et la profondeur de la poitrine (variable Y). Pour un échnatillon de $n = 22$ vaches, les observations sont les suivantes :

x_i	145	146	148	148	150	150	151	151	154	155	157
y_i	66	71	67	68	65	65	69	70	67	71	71
x_i	158	159	159	161	161	163	165	168	169	169	176
y_i	70	70	73	69	73	69	69	71	68	74	74

Les calculs suivants ont déjà été effectués :

$$\sum_{i=1}^{2} 2x_i = 3463, \quad \sum_{i=1}^{2} 2(x_i)^2 = 546625, \quad \sum_{i=1}^{2} 2y_i = 1530, \quad \sum_{i=1}^{2} 2(y_i)^2 = 106550, \sum_{i=1}^{2} 2x_i y_i = 241116$$

1. (facultatif) Représenter le nuage de points. Commenter.
2. Calculer, pour chacune des deux variables observées, la moyenne, la variance et l'écart-type.
3. Calculer la covariance entre ces deux variables.
4. En déduire le coefficient de corrélation linéaire empirique entre ces deux variables. Commenter.
5. Tester au niveau de signification $\alpha = 5\%$, la nullité de coefficient de corrélation linéaire.
6. Calculer les coefficients a et b de l'équation de la droite de régression linéaire de Y sur X (méthode des moindres carrés) : $Y = aX + b$.
7. (facultatif) Tracer cette droite de régression sur le graphique du nuage de point
8. Donner les meilleures prédictions linéaires de y, au sens des moindres carrées, pour les valeurs observées de X égales à 150, 157.4 et 170.
9. Donner les intervalles de confiance au niveau 95% de la valeur de Y quand X prend les valeurs 150, 157.4 et 170. Commenter.