

Fiche de Biostatistique

Tests élémentaires : Exercices

A.B. Dufour & D. Chessel

Résumé

La fiche contient quelques exercices simples sur la pratique des tests élémentaires.

Plan

1. EVENEMENTS	1
2. HYPOTHESES.....	2
3. QUELLE INFORMATION CONTIENNENT CES DONNEES ?.....	3
4. EXAMEN JUIN 2000 (1 HEURE).....	9
5. EXAMEN SEPTEMBRE 2000 (1 HEURE).....	13

1. Evènements

Génotypes - On considère le croisement de deux animaux de génotypes Aa et on suppose que l'allèle a est récessif. On désigne par [A] et [a] les phénotypes correspondants. Quelles sont les probabilités d'obtenir dans la F_1 un individu [A], un individu [a]? Quelle est la probabilité qu'un individu [A] de la F_1 soit Aa ? On croise deux individus [A] et [a] de la F_1 . Quelle est la probabilité qu'un de leurs descendants soit [a] ?

Temps d'attente - Dans un groupe de 10 souris, deux d'entre elles sont atteintes d'une certaine maladie (non contagieuse). On les identifie en testant les souris l'une après l'autre. Le but est de trouver les deux souris malades. Calculer la probabilité pour que le test :
soit terminé au bout de deux tirages
nécessite plus de trois tirages.

Anniversaire - Quelle est la probabilité que deux étudiants au moins dans votre groupe de TD aient même date anniversaire (quel que soit l'âge) ?

Dépistage - Un test de dépistage d'une maladie m est positif avec 94% de certitude et négatif avec 92% de certitude. Supposons que le résultat de mon test soit positif, sachant qu'à mon âge, une personne sur 120 a la maladie m , quelle est la probabilité que j'aie effectivement cette maladie ?

Probabilité conditionnelle - Dans une population d'insectes, le nombre de mâles est le tiers de celui des femelles et la fréquence du caractère ailes vestigiales est 1% chez les mâles et 0.75% chez les femelles. Quelle est la probabilité :

qu'un insecte pris au hasard ait des ailes vestigiales ?

qu'un insecte ayant des ailes vestigiales soit mâle ?

Loi hypergéométrique - (*in Probabilités et Statistiques pour biologistes*. F. Couty, J. Debord et D. Fredon, 1990, Armand Colin)

Un agriculteur a entreposé dans un local humide 12 doses d'un herbicide total et 8 doses d'un fongicide. Après plusieurs mois de séjour, les étiquettes sont indifférenciables. Chaque dose a la même probabilité d'être tirée. En vue d'un traitement, l'agriculteur prend 6 doses au hasard. Soit X la variable aléatoire égale au nombre de doses d'herbicide prises parmi ces 6 doses. Déterminer la loi de probabilité de X . Donner la moyenne et la variance de cette distribution.

Dés - Pour fournir sa note de fin d'année, un professeur demande à l'étudiant de lancer deux dés. Il considère ensuite la plus grande des deux valeurs obtenues. Il définit alors une variable aléatoire N appelée note, égale à trois fois la valeur conservée après le jet des deux dés. Déterminer la loi de probabilité de N . Donner la moyenne et la variance de cette distribution.

Loi binomiale - Le taux de réussite en DEUG Science de la Vie est de 40%. Cette année, 1000 étudiants sont inscrits. Quel est, au risque 0.05, l'intervalle du nombre de reçus probables ? Même question pour les 40 étudiants de la faculté catholique.

La proportion de fumeurs est de 0.35 chez les étudiants de licence. On considère 120 étudiants et l'on parie que le nombre de fumeurs parmi eux sera compris entre 30 et 54. Quel est le risque de perdre le pari ?

2. Hypothèses

QCM - Un professeur souhaite tester l'hypothèse suivante H_0 : « tel étudiant ne connaît aucune partie du cours ». Il fait passer un QCM avec quatre questions. Pour chaque question, quatre réponses sont proposées dont une seule est correcte. On note X le nombre de réponses correctes fournies par un étudiant. On décide que si $X \geq 3$, on rejette H_0 .

On considère trois étudiants : A ne sait rien ; B connaît un tiers du cours ; B sait tout.

1) Calculer dans les différents cas α ou β .

2) Le professeur décide de changer sa décision. Il rejette H_0 si $X \geq 1$. Même question.

3) Le professeur décide de poser non plus 4 questions mais 100 questions et accepte $\alpha = 0,05$. Quelle est la zone de rejet ? Que vaut β pour l'étudiant B ?

Maïs - En T.P. de génétique, on étudie un épi de maïs à grains noir ou jaune, par binômes. A partir des comptages des grains des deux couleurs, proposer un modèle mendélien pouvant rendre compte des résultats observés : $J=265$, $N=235$. L'étudiant 1 propose :

la couleur vient d'un gène biallélique : $[A] \rightarrow J$ $[a] \rightarrow N$

l'épi est issu d'un croisement $Aa \times aa$

L'étudiant 2 propose :

la couleur vient de deux gènes bialléliques : $[AB] \rightarrow J$ sinon $\rightarrow N$

l'épi provient d'un croisement $AaBb \times AaBb$

l'allèle A domine l'allèle a ; l'allèle B domine l'allèle b

Comparer la vraisemblance de ces deux hypothèses.

Loi exponentielle - On suppose que X suit la loi exponentielle de paramètre q : $f(x, q) = qe^{-qx}$ avec $x > 0$. On souhaite tester $H_0 : q = 2$ contre $H_1 : q = 1$ sur la base d'une seule observation et avec comme zone de rejet de $H_0 : x \geq 1$. Calculer les risques de première et de deuxième espèces associés à ce test.

3. Quelle information contiennent ces données ?

Corégone - (in « Biostatistique » B. Scherrer, 1984, Gaëtan Morin éditeur.) Dans le cadre d'une étude sur le cycle vital du grand corégone (*Coregonus clupeaformis*) du lac Nathalie, situé dans le territoire de la baie James, Dumont (1977) a mesuré la longueur totale du corps (en mm) de 756 individus.

275	285	295	305	315	325	335	345	355	365	375
1	1	4	2	5	2	7	6	11	17	13
385	395	405	415	425	435	445	455	465	475	485
25	38	72	102	140	107	77	52	34	21	4
495	505	515	525	535	545	555	565	575	585	595
6	4	2	0	0	1	0	1	0	0	1

Tribolium - Pour un échantillon de 20 *tribolium confusum* mâles, on a noté la distance parcourue, en mm, pendant une minute. Les résultats sont rangés par ordre croissant :

100	116	136	138	140	150	155	160	166	170
192	218	227	238	243	250	252	263	277	280

Guérison - Sur deux groupes de malades, on essaie deux thérapies. Dans le premier groupe, constitué de 200 individus, on observe 72% de guérisons et dans le second groupe, constitué de 100 individus, on observe 88% de guérisons.

Anesthésie - (in « Introduction aux biostatistiques » S.A. Glantz, 1998, Mc Graw-Hill.)

Le type d'anesthésie (halothane ou morphine) et l'issue de l'intervention ont été enregistrés dans 128 opérations opération à cœur ouvert .

Anesthésie	Vivants	Décédés
Halothane	53	8
Morphine	57	10

Vaccin - Pour trois modes de préparation d'un vaccin (A,B,C) on mesure la réaction locale au point d'injection chez 500 patients. On obtient le tableau ci-dessous.

Réaction	Légère	Moyenne	Ulcération	Abcès
Vaccin A	13	158	8	1
Vaccin B	30	133	6	1
Vaccin C	9	129	10	2

Bruit - On fait passer un test de réactivité visuelle à un groupe de 118 sujets. Chaque sujet passe le test dans deux conditions différentes : avec ou sans bruit dans la salle. Le test se présente comme suit. Deux lampes sont placées à droite et à gauche du sujet. Chaque lampe s'allume de façon aléatoire, avec un temps d'attente variable (entre 0.2s et 0.5s). Le sujet est assis les mains sur les genoux. Dès qu'une lampe s'allume, il doit frapper une plaque située en dessous de la lampe correspondante. On considère qu'un sujet a réussi le test lorsqu'il a réalisé la bonne association « lumière, frappe » au moins 7 fois sur 10.

	<i>avec bruit</i>	
<i>sans bruit</i>	<i>Succès</i>	<i>Echec</i>
<i>Succès</i>	62	26
<i>Echec</i>	7	23

Latéralité - On étudie la latéralité chez 56 étudiants de l'UFR STAPS. Le caractère gaucher ou droitier est défini classiquement selon la main d'écriture. Les étudiants sont divisés en deux groupes : les joueurs de tennis et les autres. Les données sont consignées dans le tableau ci-dessous.

	<i>Droitier</i>	<i>Gaucher</i>
<i>Joueurs de tennis</i>	11	3
<i>Non joueurs de tennis</i>	34	8

Test de Fisher - Pour savoir si les chercheurs se rendent compte ou non s'ils utilisent ou non un test exact de Fisher unilatéral ou bilatéral, W.P. McKinney et ses collaborateurs (in " The inexact Use of Fisher's exact test in six major medical journals ", McKinney, Young, Harta & Lee, *JAMA*, 1989, 261 ; 3430-3433) ont examiné l'utilisation du test exact de Fisher dans des articles de deux grandes revues médicales (*New England Journal of Medicine* (NEJM) et *Lancet*) afin de vérifier si les auteurs notaient le type de test exact de Fisher utilisé (Oui-Non).

	Oui	non
NEJM	1	8
Lancet	10	4

Diagnostique - (in « Biostatistique » B. Scherrer, 1984, Gaëtan Morin éditeur)

Le rétrécissement des artères et des veines sous claviaires au niveau de l'articulation du bras engendre chez des patients des démangeaisons pouvant nécessiter des interruptions de travail. Le diagnostic du syndrome peut être posé grâce à l'angiographie (c'est-à-dire la radiographie des vaisseaux après injection d'un liquide opaque aux rayons X) effectuée sur des patients en position assise ou couchée. Pour tester la position la plus efficace, on a relevé la présence (positif) ou l'absence (négatif) de détection de la maladie chez 112 patients.

	Position	effectif
	<i>assise</i> <i>couchée</i>	
positif	positif	59
positif	négatif	8
négatif	positif	20
négatif	négatif	25

Balsamine (in *Probabilités et Statistiques pour biologistes* F. Couty, J. Debord et D. Fredon, 1990, Armand Colin.) On a effectué le croisement de balsamines blanches avec des balsamines pourpres. En première génération, les fleurs sont toutes pourpres. On obtient en deuxième génération quatre catégories avec les effectifs suivants :

Couleurs	pourpre	rose	blanc lavande	blanc
Effectifs	1790	547	548	213

Peut-on accepter l'hypothèse de répartition mendélienne $\left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$?

Filtres in « Biostatistique » B. Scherrer, 1984, Gaëtan Morin éditeur. Dans une étude sur le traitement des eaux usées, l'efficacité de deux filtres, l'un en fibre de verre et l'autre en papier filtre Whatman n°40 a été testée. Sur des prélèvements de 200 millilitres d'eau provenant d'une usine de pâtes à papier, la quantité de solides en suspension retenus par les deux filtres a été mesurée. L'efficacité du filtre de verre est-elle supérieure à celle du papier filtre ?

N° prélèvement	Fibre de verre	Papier
1	68	64
2	86	75
3	91	72
4	77	63
5	100	96
6	138	120
7	190	184
8	110	96

Alcool in « Applied Statistics for the Behavioral Sciences », D.E. Honkle, W. Wiersma, S.G. Jurs, 1988, Houghton Mifflin.) On veut étudier le temps de réaction nécessaire pour arrêter une automobile chez des sujets sous l'influence de trois onces d'alcool (environ 0,09 l). On mesure le temps de réaction (en 100^{ème} de seconde) avant et après l'ingestion d'alcool.

<i>avec alcool</i>	<i>sans alcool</i>
46	33
41	29
37	26
37	23
30	21
43	36
38	27
47	38
33	22
42	33
54	42
48	35
33	22
54	39
50	37

Silex - On a mesuré la dureté du 12 silex provenant de deux régions A et B, et on a classé les silex du moins dur au plus dur.

Rang	1	2	3	4	5	6	7	8	9	10	11	12
Origine	A	A	A	B	A	A	B	A	B	B	B	B

La dureté est-elle la même dans les deux régions ?

Fécondité - V. Labeyrie a étudié la fécondité d'une guêpe parasite (*Diadromus pulchellus*) en fonction de la présence d'un hôte au stade cocon (*Acrolepija assectella*). Deux lots A et B de parasites sont formés. Aux parasites du lot A, on présente un hôte tous les jours (du 1^{er} au 35^{ème} jour), à ceux du lot B à partir du 6^{ème} jour seulement (du 6^{ème} au 35^{ème} jour). Les nombre d'œufs pondus par chaque insecte sont les suivantes :

Lot A (13 insectes) :

98	84	63	75	84	66	56	48	48	109	85	95	106
----	----	----	----	----	----	----	----	----	-----	----	----	-----

Lot B (18 insectes) :

124	83	75	123	105	108	155	128	56	72	96	45	71	45	73	60	89	83
-----	----	----	-----	-----	-----	-----	-----	----	----	----	----	----	----	----	----	----	----

Lierre - On a mesuré sur 10 feuilles de lierre la longueur de la nervure principale et la longueur du pétiole (en mm).

Nervure	76.5	70.5	81.5	82	67.5	86	77.5	66.5	64	67
Pétiole	27	34.5	33.5	37	32.5	42	41	44	38	35.5

Math-Physique - Les notes obtenues par 20 étudiants lors d'un examen ont été :

<i>Math</i>	12	18	2	30	12	8	22	26	10	32
<i>Physique</i>	14	20	0	24	6	18	18	24	14	34
<i>Math</i>	16	28	10	6	4	14	18	14	8	10
<i>Physique</i>	20	6	14	16	14	0	18	20	4	6

Rats

Milieu	<i>Poids des rats (g)</i>						
A	257	205	206	190	214	228	203
B	201	231	197	185			
C	248	265	187	220	212	215	281
D	202	276	207	230	204	227	

On a pesé, à l'âge de 12 semaines, des rats élevés sur 4 milieux nutritifs différents (A,B,C,D).

Cultures - Quatre méthodes de culture de maïs ont conduit aux rendements ci-dessous.

1	83	91	94	89	89	96	91	92	90	
2	91	90	81	83	84	83	88	91	89	84
3	101	100	91	93	96	95	94			
4	78	82	81	77	79	81	80	81		

Café - Douze goûteurs de café doivent choisir entre quatre essences et les classer. Le résultat est le suivant :

	1	2	3	4	5	6	7	8	9	10	11	12
<i>Essence A</i>	4	4	3	3	4	3	1	2	3	4	4	3
<i>Essence B</i>	3	2	1	1	2	1	3	4	1	1	2	1
<i>Essence C</i>	2	3	2	2	1	2	2	1	2	3	3	2
<i>Essence D</i>	1	1	4	4	3	4	4	3	4	2	1	4

Anesthésie - On anesthésie des larves de *Bombyx mori* et on mesure leur temps de réveil (en minutes) en fonction de la température dans l'élevage et de la durée de l'anesthésie (D. Sillans). Le résultat est le suivant :

<i>Durée</i>	<i>Température</i>		
	20°C	25°C	30°C
10 mn	4	5.5	5
20 mn	6	7	7
30 mn	12	17	13
40 mn	17	23	21

Bibliothèque - in *Théorie et Applications de la Statistique* M.R. Spiegel, 1972, Mc Graw-Hill. Le tableau ci-dessous indique le nombre de livres empruntés dans une bibliothèque municipale pendant une semaine déterminée.

	Lundi	Mardi	Mercredi	Jeudi	Vendredi
Nombre	135	108	120	114	146

Force - On mesure la force statique par dynamométrie manuelle (exprimée en kg) chez 10 enfants Trisomique 21. Un premier test est réalisé en septembre. Puis, sur une période de six mois, ces enfants essaient de développer, sous forme de jeux, leur force statique. Un second test est réalisé en février. Peut-on dire que le programme suivi par ces enfants a permis d'améliorer leur résultat ?

33	42,5	54	60	61	68	68	69	72	86
38	45	52	63	61	75	66	70	81	90

Désastres - in Cox, D.R. & Lewis, P.A.W. (1969) *L'analyse statistique des séries d'évènements*. Traduction de Larrieu (J.) Dunod, Paris. p. 4 d'après Maguire, B.A., Pearson, E.S. & Wynn, A.H.A. (1952) The time intervals between industrial accidents. *Biometrika* : 39, 168-180.)

<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>		<i>n</i>	
1	0	16	1507	31	3374	46	5231	61	10652	76	16051	91	20872	106	24664
2	378	17	1683	32	3482	47	5382	62	11150	77	16105	92	20891	107	24881
3	414	18	1738	33	3670	48	5743	63	11199	78	16431	93	21220	108	24888
4	429	19	1831	34	3903	49	6055	64	11330	79	17743	94	21550	109	24906
5	460	20	1890	35	3931	50	6409	65	11512	80	18091	95	21862	110	26263
6	675	21	2205	36	3953	51	6467	66	11767	81	18836	96	22033		
7	686	22	2264	37	4014	52	6742	67	11962	82	19053	97	22178		
8	823	23	2325	38	4092	53	6820	68	12186	83	19173	98	22253		
9	827	24	2326	39	4191	54	6837	69	12752	84	19448	99	22617		
10	842	25	2339	40	4517	55	8042	70	13142	85	19468	100	22654		
11	914	26	2528	41	4792	56	8686	71	13214	86	19534	101	22673		
12	1010	27	2873	42	4846	57	9153	72	13442	87	19825	102	22829		
13	1134	28	2893	43	5063	58	10024	73	13713	88	19829	103	22876		
14	1184	29	2974	44	5176	59	10072	74	13921	89	20198	104	23005		
15	1304	30	3260	45	5208	60	10195	75	14438	90	20536	105	24635		

La statistique donne la date en jours de 110 désastres survenus dans les mines de charbon en Grande-Bretagne pendant la période 1875-1951. Un désastre implique la mort de dix hommes ou plus.

4. Examen Juin 2000 (1 heure)

Question 1. Quelle est la valeur qui a une chance sur 100 d'être dépassée par une variable aléatoire qui suit une loi Khi2 à 5 degrés de liberté ?

Question 2. Quelle est la valeur de la quantité $\frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-\frac{t^2}{2}} dt$?

Question 3. Vous jouez contre un inconnu à un jeu de hasard 11 fois. Vous gagnez 2 fois. Dites vous :

A - « Bof, c'est le hasard »

B - « Il est en train de me rouler » ?

a est un échantillon de 10 valeurs d'une loi normale de moyenne -0.2 et de variance 1. b est un échantillon de 20 valeurs d'une loi normale de moyenne 0.2 et de variance 1.

> a

```
[1] -0.6950  0.9563  0.3371 -1.2093  1.0423 -0.3486 -0.2193  0.6161 -0.5576
```

```
[10]  1.1069
```

> b

```
[1]  1.84556  0.53173  0.24699  1.10628  0.19009 -0.55104 -1.73625 -1.60875
```

```
[9]  1.27278  1.97929  0.17162 -0.03924  0.38513  1.29323  1.24422  1.00132
```

```
[17]  1.91226  0.18235 -0.61401  0.87469
```

Le test t de comparaison de ces deux échantillons donne :

```
data:  a and b
```

```
t = -1.091, df = 22.65, p-value = 0.2869
```

Le test de Wilcoxon-Mann-Whitney de comparaison de ces deux échantillons donne :

```
data:  a and b
```

```
W = 74, p-value = 0.267
```

Question 4. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

A - Ces résultats sont aberrants parce que les deux moyennes théoriques sont différentes.

B - Ces résultats sont aberrants parce que les deux tests sont différents.

C - Ces résultats sont aberrants parce que les deux effectifs sont différents.

D - Ces résultats ne sont pas aberrants.

Je veux tester par un test de Wilcoxon l'hypothèse nulle « Les deux échantillons sont extraits de la même population » contre l'hypothèse alternative « La moyenne de l'échantillon 1 est plus faible que celle de l'échantillon 2 ». Je dispose de 3 observations pour l'échantillon 1 et de 4 observations pour l'échantillon 2. Je ne dispose pas de la table de la distribution de Wilcoxon.

Question 5. Ce test peut-il être significatif au seuil de 5% ?

On a enregistré le nombre de jours qui séparent deux tremblements de terre graves sur la surface de la terre. Un tremblement de terre est grave si sa magnitude est au moins égale à 7.5 sur l'échelle de Richter ou s'il a causé la mort d'au moins 1000 personnes. L'enregistrement couvre une période de 27300 jours entre 1902 et 1977. Il y a eu 63 tremblements de terre et donc 62 valeurs du temps d'attente du suivant.

```
> delai Temps d'attente du suivant
[1] 840 157 145 44 33 121 150 280 434 736 584 887 263 1901 695
[16] 294 562 721 76 710 46 402 194 759 319 460 40 1336 335 1354
[31] 454 36 667 40 556 99 304 375 567 139 780 203 436 30 384
[46] 129 9 209 599 83 832 328 246 1617 638 937 735 38 365 92
[61] 82 220
> dat Date de la catastrophe
[1] 100 940 1097 1242 1286 1319 1440 1590 1870 2304 3040 3624
[13] 4511 4774 6675 7370 7664 8226 8947 9023 9733 9779 10181 10375
[25] 11134 11453 11913 11953 13289 13624 14978 15432 15468 16135 16175 16731
[37] 16830 17134 17509 18076 18215 18995 19198 19634 19664 20048 20177 20186
[49] 20395 20994 21077 21909 22237 22483 24100 24738 25675 26410 26448 26813
[61] 26905 26987 27207
```

[Source : Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. & Ostrowski, E. (1994) A handbook of small data sets. Chapman & Hall, London. 1-458. p. 203.]

La moyenne de la statistique delai vaut 437.21. On fait l'hypothèse que le temps d'attente d'une catastrophe aléatoire suit une loi définie par $P(X < t) = \int_0^t a e^{-ax} dx = 1 - e^{-at}$ (loi exponentielle).

Question 6. Quelle est l'estimation de a au maximum de vraisemblance ?

Les valeurs observées de delai sont rangées en 5 classes. Les fréquences, les probabilités et les effectifs théoriques associés à la loi exponentielle sont calculés.

Classes]0,100]]100,200]]200,400]]400,800]	>800	Total
Effectifs observés	14	7	14	19	8	62
Fréquences	0.226	0.113	0.226	0.306	0.129	1
Probabilités	0.204	0.163	0.232	0.240	0.160	1
Effectifs théoriques	12.7	10.1	14.4	14.9	9.9	62
(Obs-The)^2/The	0.138	0.943	0.011	1.136	0.381	2.611

Question 7. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

- A - La variable étudiée suit une loi exponentielle.
- B - La variable étudiée ne suit pas une loi exponentielle.
- C - On ne peut pas affirmer que la variable étudiée ne suit pas une loi exponentielle.
- D - La probabilité pour que la variable étudiée suive une loi exponentielle est faible.

On peut tester l'ajustement par le test de Kolmogorov-Smirnov sur la loi exponentielle :

```
> ks.test(delai, "pexp", 1/mean(delai))

One-sample Kolmogorov-Smirnov test

data: delai
```

D = 0.0745, p-value = 0.8816
 alternative hypothesis: two.sided

Question 8. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

- A - La variable étudiée suit une loi exponentielle.
- B - La variable étudiée ne suit pas une loi exponentielle.
- C - On ne peut pas affirmer que la variable étudiée ne suit pas une loi exponentielle.
- D - La probabilité pour que la variable étudiée suive une loi exponentielle est faible.

La période d'étude est divisée en 100 intervalles et on compte le nombre de tremblements de terre par intervalle. On trouve :

```
> n100
[1] 1 0 0 1 4 2 1 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 2 0
[38] 1 1 0 1 1 0 2 0 0 0 0 1 1 0 0 0 0 1 0 2 0 0 2 0 2 1 0 1 0 2 0 0 1 1 1 1 3
[75] 1 0 1 1 0 0 1 1 1 0 0 0 0 0 1 0 1 0 0 0 1 0 2 0 3 1
> mean(n100)
[1] 0.63
> table(n100)
n100
 0  1  2  3  4
52 37  8  2  1
```

Catastrophes	0	1	>1	Total
Effectifs observés	52	37	11	100
Fréquences	0.520	0.370	0.110	1
Probabilités	0.532	0.336	0.132	1
Effectifs théoriques	53.2	33.6	13.2	100
$(\text{Obs}-\text{The})^2/\text{The}$	0.027	0.344	0.367	0.738

Question 9. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

- A - La valeur du Khi2 est significativement petite.
- B - La valeur du Khi2 est significativement grande.
- C - La valeur du Khi2 n'est pas significative.
- D - Aucune des assertions qui précèdent n'a un sens.

La période d'étude est divisée en 10 intervalles et on compte le nombre de tremblements de terre par intervalle. On trouve :

```
> break1_seq(0,27300,le=11)
> n10_hist(dat,breaks=break1,plot=F)$counts
> n10_hist(dat,breaks=break1,plot=F)$counts
> n10
[1] 10 4 3 7 6 5 7 9 4 8
> mean(n10)
[1] 6.3
> > sum((n10-mean(n10))^2/mean(n10))
[1] 7.63492
```

Intervalle	1	2	3	4	5	6	7	8	9	10	Total
Effectifs observés	10	4	3	7	6	5	7	9	4	8	63
Fréquences	0.16	0.06	0.05	0.11	0.10	0.08	0.11	0.14	0.06	0.13	1
Probabilités	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	1
Effectifs théoriques	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	63
$(\text{Obs}-\text{The})^2/\text{The}$	2.17	0.84	1.73	0.08	0.01	0.27	0.08	1.16	0.84	0.46	7.635

Question 10. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

- A - Le test est fortement significatif.
- B - Le test est significatif.
- C - Le test n'est pas significatif.
- D - Aucune des assertions qui précèdent n'a un sens.

On peut tester l'ajustement par le test de Kolmogorov-Smirnov sur la loi uniforme de la variable dat :

```
> ks.test(dat, "punif", 0, 27300)
```

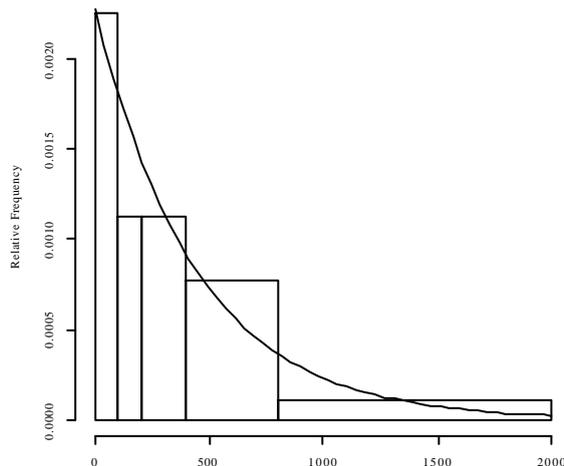
```
One-sample Kolmogorov-Smirnov test
```

```
data: dat
D = 0.0744, p-value = 0.877
alternative hypothesis: two.sided
```

Question 11. Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

- A - La série des tremblements de terre ne peut être un processus aléatoire.
- B - La série des tremblements de terre n'est pas un processus aléatoire.
- C - La série des tremblements de terre peut être un processus aléatoire.
- D - La série des tremblements de terre est un processus aléatoire.

Question 12. Rédiger une légende pour la figure donnée.



SOLUTION

Question 1

La réponse est 15.1. Table 5 La loi du χ^2 du cours, ligne 5, colonne 0.99

Question 2

Réponse Table 2 La loi normale $p=1-2P(X<=-1)=1-2*0.1587=0.6826$

Question 3

Loi binomiale $p=0.5$ $n= 11$ 0.033 dans la table du cours

Je dis « Il est en train de me rouler avec une probabilité de 3/100 de me tromper si le jeu est honnête ».

Question 4

L'hypothèse nulle est fautive et les tests ne sont pas significatifs. Ce n'est pas impossible à cause du risque de deuxième espèce. Je choisis D.

Question 5

Si les trois valeurs de l'échantillon 1 sont inférieures aux quatre valeurs de l'échantillon 2, la somme des rangs vaut 6. La probabilité pour qu'elle soit inférieure ou égale à 6 vaut $3!/7! = 1/35$. Le test sera significatif à 5%.

Question 6

Chapitre 2 p. 14 $1/m(x)=1/437.21=0.00229$.

Question 7

Le χ^2 d'ajustement 2.61 pour $ddl=5-1-1=3$ n'est pas significatif. Je ne rejette pas l'hypothèse nulle. Je choisis C.

Question 8

La probabilité de trouver un écart plus grand entre le modèle et les données vaut 0.88. Je ne rejette pas l'hypothèse nulle. Je choisis C.

Question 9

Le χ^2 d'ajustement 0.74 pour $ddl=3-1-1=1$ n'est pas significatif. Le seuil à gauche est 0.0 à 5%. Le seuil à droite est 3.8 à 5%. Je choisis C.

Question 10

Le χ^2 d'ajustement 7.63 pour $ddl=10-1=9$ n'est pas significatif. Le seuil à droite est 15.5 à 5%. Je choisis C.

Question 11

Aucun des tests proposés n'est significatif. Je n'ai aucun argument pour rejeter l'hypothèse d'un processus aléatoire. Mais on ne peut pas accepter une hypothèse nulle. Je choisis C.

Question 12

Histogramme de la statistique délai de 62 temps d'attente d'un tremblement de terre. Superposition de la densité exponentielle de paramètre 0.00229 (Question 7). Les fréquences des classes sont représentées par la surface d'un rectangle et les probabilités par l'aire sous la courbe. L'écart entre théorique et observé n'est pas significatif (Question 8 et 9).

5. Examen Septembre 2000 (1 heure)

Question 1 Quelle est la probabilité p pour qu'une variable aléatoire qui suit une loi du χ^2 à 5 degrés de liberté dépasse la valeur 11.1 ?

Quelle est la valeur x qui a la probabilité 0.05 d'être dépassée par une variable aléatoire qui suit une loi du Khi2 à 5 degrés de liberté ?

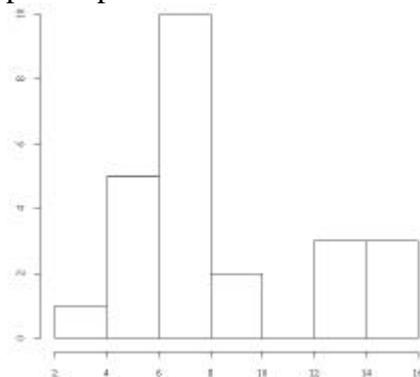
Question 2 Quelle est la valeur de la quantité $\frac{1}{\sqrt{2p}} \int_{-1}^1 e^{-\frac{t^2}{2}} dt$?

Question 3 Une urne contient 12 boules blanches et 12 boules noires. Quelle est la probabilité pour qu'un tirage sans remise contienne entre 0 et 3 boules noires ?

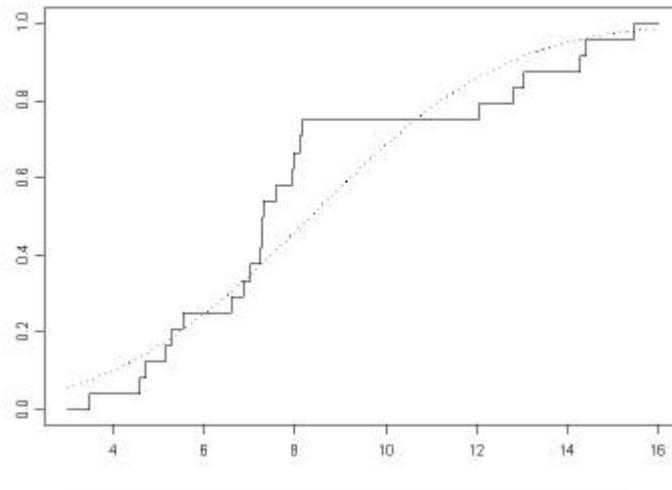
La caféine agit sur les systèmes cardio-vasculaires et nerveux et sur divers processus métaboliques. Les femmes enceintes mettent plus de temps à éliminer la caféine de leur corps que lorsqu'elles ne sont pas enceintes. Il est possible que les variations d'équilibre hormonal durant la grossesse modifient la manière dont la caféine est métabolisée en ralentissant son élimination. Pour étudier l'existence de différences du métabolisme de la caféine chez la femme, on a administré une pilule de 250mg de caféine (environ trois tasses de café) à des femmes qui prenaient des contraceptifs oraux (oui pour CO) et des femmes qui n'en prenaient pas (non pour CO). Les femmes prenant des contraceptifs oraux ont un métabolisme semblable à celui des femmes enceintes. On a mesuré les concentrations de caféine et mesuré la vitesse avec laquelle la caféine disparaissait dans le sang (taux de demi-vie c'est-à-dire le temps que prend la concentration de caféine pour arriver à la moitié de sa valeur initiale, exprimé en heures). Les résultats obtenus sont :

	1	2	3	4	5	6	7	8	9	10	11	12
vitesse	3.47	4.59	4.72	5.17	5.3	5.54	6.59	6.87	7.01	7.25	7.26	7.28
CO	non	non	non	non	non	oui	non	oui	non	non	oui	non
	13	14	15	16	17	18	19	20	21	22	23	24
vitesse	7.3	7.6	7.94	7.98	8.11	8.16	12.04	12.81	13.04	14.28	14.41	15.47
CO	non	non	oui	oui	oui	non	oui	oui	oui	oui	oui	oui

Question 4 Compléter la figure pour exprimer la variable CO et donner une légende :



Question 5 Que représente la figure ?



Question 6 Quel est le test pratiqué ci-dessous et que signifie le résultat ? Quelles vérifications sont elles nécessaires pour valider son usage ?

```
> oui
[1] 5.54 6.87 7.26 7.94 7.98 8.11 12.04 12.81 13.04 14.28 14.41 15.47
> non
[1] 3.47 4.59 4.72 5.17 5.30 6.59 7.01 7.25 7.28 7.30 7.60 8.16
> t.test(oui,non)

data: oui and non
t = 3.89, df = 22, p-value = 0.0008
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.996 6.556
sample estimates:
mean of x mean of y
 10.48    6.203
```

Question 7 Quel est le test pratiqué ci-dessous et que signifie le résultat ?

```
> wilcox.test(oui,non)

data: oui and non
rank-sum statistic W = 202, n = 12, m = 12, p-value = 0.0018
alternative hypothesis: true mu is not equal to 0
```

Question 8 On réduit les données sous la forme :

	oui	non
vitesse < 7.29	9	3
vitesse > 7.29	3	9

Donner un test de comparaison des deux échantillons basé sur cette réduction et la loi du Khi2. Interpréter le résultat.

Question 9 Utiliser le résultat de la question 3 pour donner un autre test dans les mêmes conditions. Le résultat obtenu est-il compatible avec le précédent ?

Question 10 Parmi les choix suivants, quelle est l'assertion qui vous paraît convenir ?

A - Le temps d'élimination de la caféine est significativement augmenté par la contraception orale.

B - Le temps d'élimination de la caféine est significativement diminué par la contraception orale.

C - Le temps d'élimination de la caféine n'est pas modifié significativement par la contraception orale.

D - On ne peut choisir aucune des assertions précédentes.

Question 11 Une maladie affecte 1% des individus d'une population. Le test de dépistage utilisé pour un individu malade est positif dans 80% des cas. Le test de dépistage utilisé pour un individu sain est positif dans 10% des cas. Quelle est la probabilité pour qu'un individu soit effectivement malade sachant qu'il a eu un test positif ?

Question 12 Une maladie affecte une proportion p des individus d'une population. Le test de dépistage utilisé est sans erreur. Quelle est la vraisemblance de l'hypothèse «la maladie affecte 1% des individus de cette population» sachant qu'un échantillon aléatoire de 10 individus a fourni un seul test positif ? Que peut-on dire de la vraisemblance de l'hypothèse « la maladie affecte une proportion p des individus de cette population » pour une valeur de p différente de 1% ?

Solution

Question 1 Quelle est la probabilité p pour qu'une variable aléatoire qui suit une loi Khi2 à 5 degrés de liberté dépasse la valeur 11.1 ?

Réponse 5/100 Table 5 La loi du Khi2 du cours, ligne 5, colonne 0.95

Quelle est la valeur x qui a la probabilité 0.05 d'être dépassée par une variable aléatoire qui suit une loi Khi2 à 5 degrés de liberté ?

Réponse 1.1 Table 5 La loi du Khi2 du cours, ligne 5, colonne 0.05

Question 2 Quelle est la valeur de la quantité $\frac{1}{\sqrt{2p}} \int_{-1}^1 e^{-\frac{t^2}{2}} dt$?

```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
> 1-2*0.1587
[1] 0.6826
```

Réponse Table 2 La loi normale $p=1-2P(X<=-1)=1-2*0.1587=0.6826$

Question 3 Une urne contient 12 boules blanches et 12 boules noires. Quelle est la probabilité pour qu'un tirage sans remise contienne entre 0 et 3 boules noires ?

$$p = \frac{\binom{12}{0}\binom{12}{12} + \binom{12}{1}\binom{12}{11} + \binom{12}{2}\binom{12}{10} + \binom{12}{3}\binom{12}{9}}{\binom{24}{12}}$$

$$p = \frac{(1*1+12*12+66*66+220*220)*12*11*10*9*8*7*6*5*4*3*2*1}{24*23*22*21*20*19*18*17*16*15*14*13}$$

$p = 0.01956$
`> phyper(3,12,12,12)`
`[1] 0.01956`

Question 6 Quel est le test pratiqué ci-dessous et que signifie le résultat ? Quelles vérifications sont elles nécessaires pour valider son usage ?

```
> oui
[1] 5.54 6.87 7.26 7.94 7.98 8.11 12.04 12.81 13.04 14.28 14.41 15.47
> non
[1] 3.47 4.59 4.72 5.17 5.30 6.59 7.01 7.25 7.28 7.30 7.60 8.16
> t.test(oui,non)

data: oui and non
t = 3.89, df = 22, p-value = 0.0008
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.996 6.556
sample estimates:
mean of x mean of y
 10.48    6.203
```

Question 7 Quel est le test pratiqué ci-dessous et que signifie le résultat ?

```
> wilcox.test(oui,non)

data: oui and non
rank-sum statistic W = 202, n = 12, m = 12, p-value = 0.0018
alternative hypothesis: true mu is not equal to 0
```

Question 8 On réduit les données sous la forme :

	oui	non
vitesse < 7.29	9	3
vitesse > 7.29	3	9

Donner un test de comparaison des deux échantillons basée sur cette réduction et la loi Khi2. Interpréter le résultat.

```
> t0
              non oui
3.00+ thru 7.29  9   3
7.29+ thru 16.00 3   9
> chisq.test(t0,correct=F)

Pearson's chi-square test without Yates' continuity correction

data: t0
```

X-square = 6, df = 1, p-value = 0.0143

```
> chisq.test(t0,correct=T)
```

Pearson's chi-square test with Yates' continuity correction

data: t0

X-square = 4.167, df = 1, p-value = 0.0412

Question 11 Une maladie affecte 1% des individus d'une population. Le test de dépistage utilisé pour un individu malade est positif dans 80% des cas. Le test de dépistage utilisé pour un individu sain est positif dans 10% des cas. Quelle est la probabilité pour qu'un individu soit effectivement malade sachant qu'il a eu un test positif ?

$$p(\text{malade} / \text{positif}) = \frac{p(\text{malade})p(\text{positif} / \text{malade})}{p(\text{malade})p(\text{positif} / \text{malade}) + p(\text{sain})p(\text{positif} / \text{sain})}$$

$$p = \frac{0.01 * 0.8}{0.01 * 0.8 + 0.99 * 0.1} = 0.075$$

Question 12 Une maladie affecte une proportion p des individus d'une population. Le test de dépistage utilisé est sans erreur. Quelle est la vraisemblance de l'hypothèse «la maladie affecte 1% des individus de cette population» sachant qu'un échantillon aléatoire de 10 individus a fourni un seul test positif ? Que peut-on dire de la vraisemblance de l'hypothèse « la maladie affecte une proportion p des individus de cette population » pour une valeur de p différente de 1%.

$L = \binom{10}{1} * 0.01^1 * 0.99^9 = 0.091$. Toutes les autres valeurs sont plus petites car l'estimation au maximum de vraisemblance de la fréquence est la probabilité.