

## Travaux pratiques III : questions

**1. Test d'adéquation en génétique.** La statistique  $\chi^2$  peut être utilisée pour tester l'hypothèse  $\mathcal{H}_0$  qu'une variable statistique  $X$  est distribuée selon une certaine distribution théorique (modèle) dans la population. Nous considérons ici le cas d'une variable  $X$  discrète dont les valeurs possibles sont  $a, b, c$ , etc. Supposons que, selon un certain modèle, les fréquences de ces valeurs dans la population sont  $p_a, p_b, p_c$ , etc. Supposons de disposer de  $n$  observations de  $X$ , que le nombre d'observations égales à  $a$  est  $n_a$ , que le nombre d'observations égales à  $b$  est  $n_b$ , que le nombre d'observations égales à  $c$  est  $n_c$ , etc. Donc  $n_a + n_b + n_c + \dots = n$ . Les "fréquences relatives observées" de  $a, b, c$ , etc sont

$$O_a = \frac{n_a}{n}, \quad O_b = \frac{n_b}{n}, \quad O_c = \frac{n_c}{n}, \quad \dots$$

Si le modèle est correct, on s'attend à observer les "fréquences relatives espérées"

$$E_a = p_a, \quad E_b = p_b, \quad E_c = p_c, \quad \dots$$

On peut alors démontrer que, si le modèle est correct ( $\mathcal{H}_0$ ) la statistique

$$\chi^2 = n \sum_{i=a,b,c,\dots} \frac{(O_i - E_i)^2}{E_i}$$

a une distribution  $\chi^2$  à  $k - 1$  degrés de libertés, où  $k$  est le nombre de modalités ("classes") de  $X$ . Une table de la distribution  $\chi^2$  est annexée.

A l'aide de ce résultat théorique et de la table, construisez un "test d'adéquation" pour le problème suivant (dihybridisme).

Un éleveur achète un couple de cobayes gris à pelage lisse. Dans les quatre ans qui suivent l'achat, l'éleveur constate que le couple de cobayes a donné naissance à 128 petits. Sur ces 128 cobayes, 78 étaient gris à pelage lisse, 19 gris à pelage rude, 26 blancs à pelage lisse et 5 blancs à pelage rude.

Les proportions de ces différents phénotypes étant proches de  $9/16, 3/16, 3/16, 1/16$ , on peut envisager, dans ce cas, un phénomène héréditaire dépendant de 2 couples d'allèles, l'un déterminant la couleur et l'autre le type du pelage. On peut supposer que le caractère gris domine le caractère blanc et que le caractère lisse domine le caractère rude. En effet, on observe en F1:

- pour le caractère (gris-blanc) :
  - $78 + 19 = 97$  gris soit environ  $3/4 \times 128$
  - $26 + 5 = 31$  blancs soit environ  $1/4 \times 128$
- pour le caractère (lisse-rude) :
  - 104 lisses soit environ  $3/4 \times 128$
  - 24 rudes soit environ  $1/4 \times 128$ .

Soit “G” (=gris) l’allèle dominant et “g” (=blanc) l’allèle récessif du couple d’allèles gouvernant la couleur du pelage. Soit “L” (=lisse) l’allèle dominant et “l” (=rude) l’allèle récessif du couple d’allèles gouvernant l’aspect du pelage.

Quel est le génotype du couple de cobayes achetés par l’éleveur ? Vérifiez votre hypothèse avec le test de Chi-carré sur les effectifs indiqués ci-dessus.

Remarque. Les tests d’adéquation ne permettent pas de trouver (ou démontrer) la “loi” d’une variable, mais seulement d’accepter ou de rejeter une hypothèse émise a priori concernant la distribution de cette variable.

**2. Analyse d’une expérience de radiophysique.** Aux pages pages 4 et 5 vous trouverez la description d’une expérience qui a pour but la comparaison de deux traitements de médecine nucléaire. Il conviendra de considérer les variables suivantes:

- $X_1$ : le No de plaquettes avant le traitement Bexxar
- $X_2$ : le No de plaquettes après le traitement Bexxar
- $Y_1$ : le No de plaquettes avant le traitement Zevalin
- $Y_2$ : le No de plaquettes après le traitement Zevalin
- $R_1$ : le rapport  $X_2/X_1$  (Bexxar)
- $R_2$ : le rapport  $Y_2/Y_1$  (Zevalin)
- $D_1$ : la diminution  $X_1 - X_2$  (Bexxar)
- $D_2$ : la diminution  $Y_1 - Y_2$  (Zevalin)

Pour vous faciliter les calculs, les moyennes empiriques et les écarts types de ces variables sont fournis dans la table suivante.

	$X_1$	$X_2$	$Y_1$	$Y_2$	$R_1$	$R_2$	$D_1$	$D_2$
moyenne	183.50	67.39	230.86	47.71	0.39	0.21	116.11	183.14
écart type	84.19	52.97	113.13	37.77	0.27	0.17	88.09	91.96

En outre, à la dernière page, vous trouverez les représentations graphiques des distributions de ces variables.

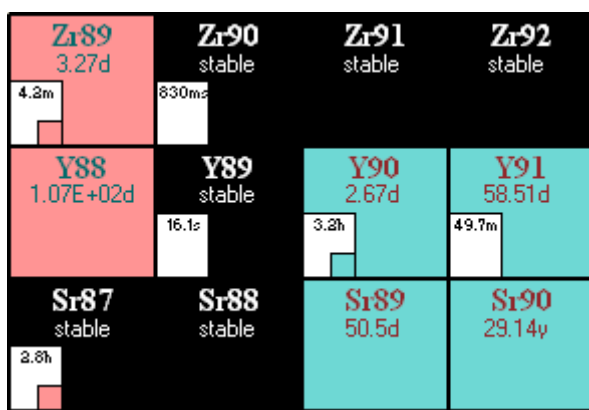
- 2.1 Sommes-nous en présence de données appariées ou de données non appariées ?
- 2.2 Quelles quantités utiliseriez-vous pour tester l'hypothèse émise à la fin de la page 4: "On suspecte que la réduction du taux de plaquettes sanguines devrait être moins importante avec le Bexxar qu'avec le Zevalin" ?
- 2.3 Quel test statistique utiliseriez-vous ? Pourquoi ?
- 2.4 Quel seuil de signification choisissez-vous ?
- 2.5 Pour le test de votre choix, formulez les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ .
- 2.6 Calculez la statistique du test et donnez une valeur approximative de la p-value (par exemple, une fourchette de valeurs). Formulez la décision à laquelle vous parvenez.
- 2.7 Voyez-vous des outliers dans les variables que vous avez utilisées pour le test ?
- 2.8 Répétez le test en supprimant l'outlier le plus marqué. Comment changent la statistique de test et la p-value ?
- 2.9 Question subsidiaire. En ce qui concerne les nombres de plaquettes avant le traitement, est-il raisonnable de supposer que les sujets traités avec Bexxar et les sujets traités avec Zevalin sont pris dans la même population ?
- 2.10 Question facultative. Pouvez-vous donner un intervalle de confiance pour la différence  $\mu(D_2) - \mu(D_1)$  entre la réduction du nombre de plaquettes sous Zevalin et la réduction du nombre de plaquettes sous Bexxar ? (La formule de cet intervalle de confiance n'a pas été donnée dans le cours; raisonnez par analogie avec l'intervalle de confiance pour une moyenne.)

## Problème de statistique basé sur un exemple de radiophysique

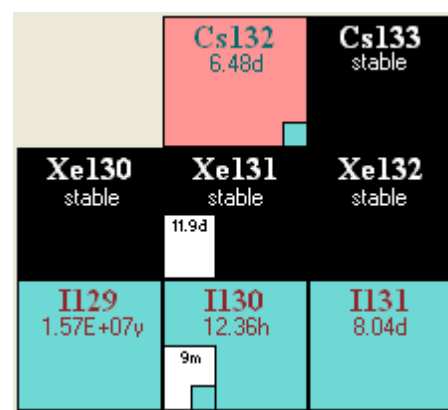
### Situation générale

En médecine nucléaire, il existe sur le marché deux anticorps monoclonaux marqués avec un isotope radioactif pour le traitement de lymphomes non-hodgkiniens (NHL) :

- Le Zevalin est un anticorps monoclonal (*ibritumomab tiuxetan*) marqué avec  $^{90}\text{Y}$ , un émetteur bêta pur.
- Le Bexxar est un autre anticorps monoclonal (*tositumomab*) marqué avec  $^{131}\text{I}$ , un émetteur bêta/gamma.



$^{90}\text{Y}$  est un émetteur bêta pur. Il n'émet rien d'autre que des particules bêta moins dont l'énergie moyenne est de 933 keV. Sa fille est le  $^{90}\text{Zr}$ , qui est stable.



$^{131}\text{I}$  est un émetteur bêta moins émettant également toute une série de photons gamma. Les particules bêta (énergie moyenne 182 keV) sont utilisées pour détruire la tumeur. Les photons gamma peuvent être utilisés pour de l'imagerie. Sa fille est le  $^{131}\text{Xe}$ , qui est stable.

Ces deux agents ont des mécanismes d'action similaires. Tous deux ciblent l'antigène CD20, qui se trouve à la surface des lymphocytes pré-B et B matures et sur plus de 90% des lymphocytes B NHL. Les anticorps reconnaissent un déterminant antigénique dans le domaine extracellulaire de l'antigène CD20 et forment un complexe anticorps-antigène qui induit l'apoptose (mort cellulaire programmée). Le radioélément ( $^{90}\text{Y}$  ou  $^{131}\text{I}$ ) contribue à ce mécanisme en émettant une particule bêta moins qui, en déposant son énergie dans la tumeur, conduit à la mort cellulaire.

Les traitements avec Zevalin et Bexxar montrent des efficacités statistiquement similaires bien que le Bexxar semble être plus efficace pour dans le cas de petites tumeurs pulmonaires (en raison du parcours plus limité des particules bêta émises par  $^{131}\text{I}$  par rapport à  $^{90}\text{Y}$ ). Dans les deux cas, le ciblage des anticorps n'est pas parfait et une partie non négligeable se retrouve dans la moelle osseuse. Ceci se traduit par une diminution du taux de plaquettes sanguines.

On suspecte que la réduction du taux de plaquettes sanguine devrait être moins importante avec l'anticorps marqué à  $^{131}\text{I}$  (Bexxar) qu'avec l'anticorps marqué à  $^{90}\text{Y}$  (Zevalin) car la dose délivrée avec le Bexxar peut être mieux ajustée à chaque patient et parce que la dose délivrée par le Bexxar tend à être plus localisée qu'avec le Zevalin.

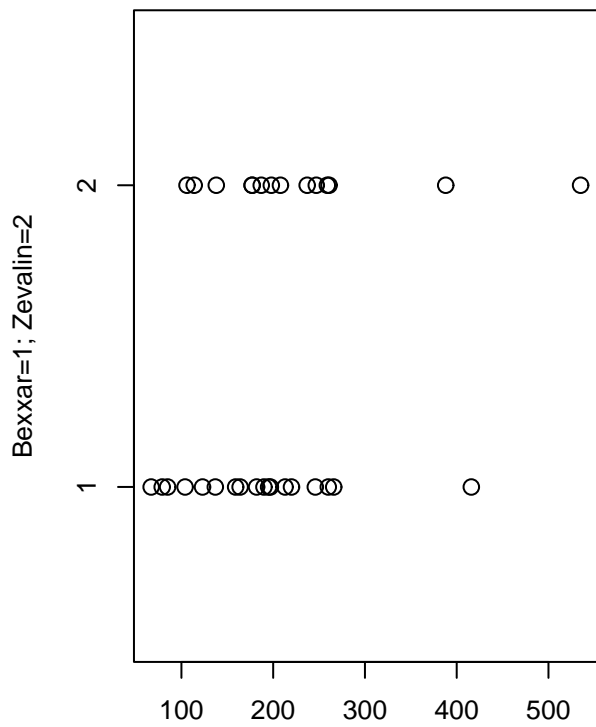
## Expérience réalisée

32 patients ont été traités pour un NHL : 18 avec du Bexxar et 14 avec du Zevalin. Le but de l'expérience était de déterminer si la réduction relative du taux de plaquettes était plus importante avec le Zevalin.

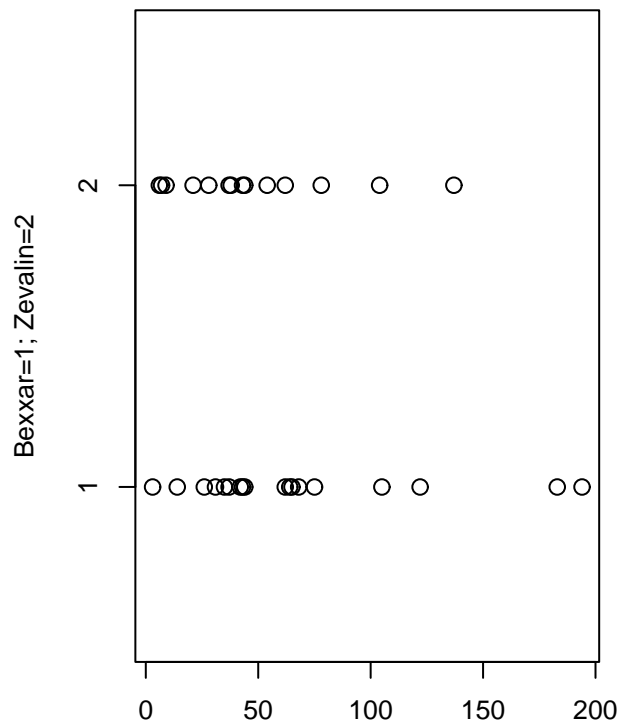
Le taux de plaquettes a été mesuré avant le traitement et après le traitement. Les résultats bruts sont donnés dans le tableau ci-dessous.

**Tableau 1 : Données brutes**

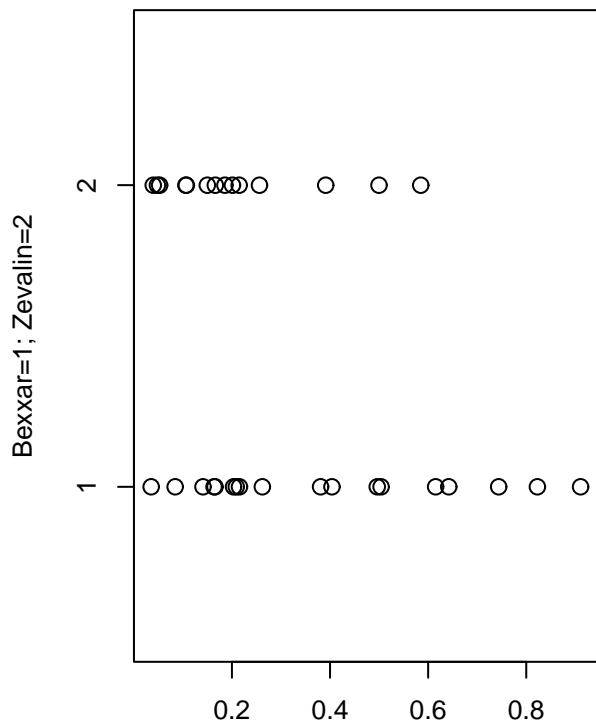
Patient n°	Anticorps	Nombre de plaquettes avant traitement ( $\times 10^{-3} \text{ mm}^{-3}$ )	Nombre de plaquettes après traitement ( $\times 10^{-3} \text{ mm}^{-3}$ )	Ratio après / avant
1	Bexxar (1.00)	197	75	0.381
2	Bexxar (1.00)	213	194	0.911
3	Bexxar (1.00)	195	42	0.215
4	Bexxar (1.00)	220	31	0.141
5	Bexxar (1.00)	416	35	0.084
6	Bexxar (1.00)	137	68	0.496
7	Bexxar (1.00)	104	64	0.615
8	Bexxar (1.00)	182	37	0.203
9	Bexxar (1.00)	79	65	0.823
10	Bexxar (1.00)	85	3	0.035
11	Bexxar (1.00)	67	14	0.209
12	Bexxar (1.00)	190	122	0.642
13	Bexxar (1.00)	260	105	0.404
14	Bexxar (1.00)	164	43	0.262
15	Bexxar (1.00)	123	62	0.504
16	Bexxar (1.00)	159	26	0.164
17	Bexxar (1.00)	246	183	0.744
18	Bexxar (1.00)	266	44	0.165
19	Zevalin (2.00)	106	62	0.585
20	Zevalin (2.00)	114	6	0.053
21	Zevalin (2.00)	208	104	0.500
22	Zevalin (2.00)	177	7	0.040
23	Zevalin (2.00)	198	21	0.106
24	Zevalin (2.00)	237	44	0.186
25	Zevalin (2.00)	535	137	0.256
26	Zevalin (2.00)	388	78	0.201
27	Zevalin (2.00)	177	38	0.215
28	Zevalin (2.00)	261	28	0.107
29	Zevalin (2.00)	138	54	0.391
30	Zevalin (2.00)	247	37	0.150
31	Zevalin (2.00)	259	43	0.166
32	Zevalin (2.00)	187	9	0.049



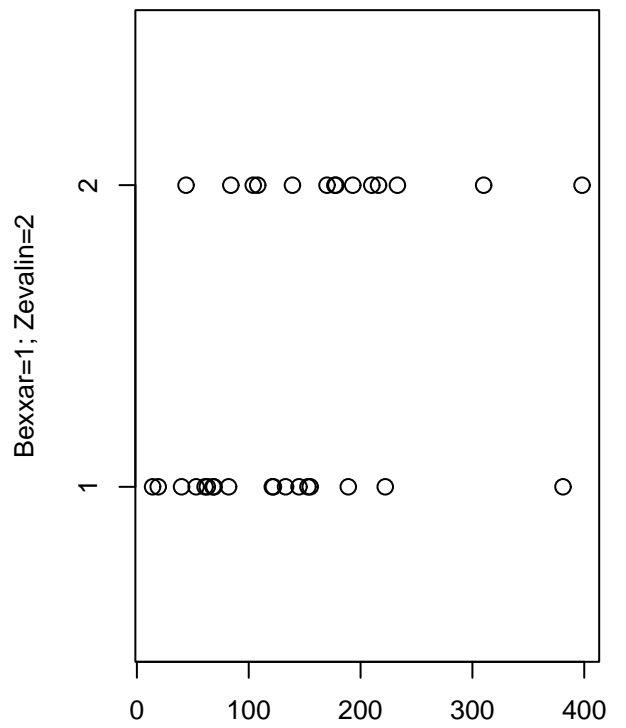
(a) No plaquettes avant traitement



(b) No plaquettes après traitement



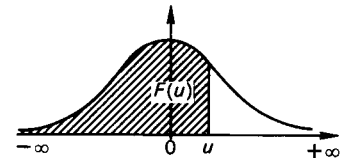
(c) Ratio après/avant



(d) Diminution avant-après

**DISTRIBUTION DE GAUSS  $\Phi(u) = F(u)$ , Fonction de distribution cumulative normale réduite (standard)**

Probabilité de trouver une valeur inférieure à u



u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 0	0,504 0	0,508 0	0,512 0	0,516 0	0,519 9	0,523 9	0,527 9	0,531 9	0,535 9
0,1	0,539 8	0,543 8	0,547 8	0,551 7	0,555 7	0,559 6	0,563 6	0,567 5	0,571 4	0,575 3
0,2	0,579 3	0,583 2	0,587 1	0,591 0	0,594 8	0,598 7	0,602 6	0,606 4	0,610 3	0,614 1
0,3	0,617 9	0,621 7	0,625 5	0,629 3	0,633 1	0,636 8	0,640 6	0,644 3	0,648 0	0,651 7
0,4	0,655 4	0,659 1	0,662 8	0,666 4	0,670 0	0,673 6	0,677 2	0,680 8	0,684 4	0,687 9
0,5	0,691 5	0,695 0	0,698 5	0,701 9	0,705 4	0,708 8	0,712 3	0,715 7	0,719 0	0,722 4
0,6	0,725 7	0,729 0	0,732 4	0,735 7	0,738 9	0,742 2	0,745 4	0,748 6	0,751 7	0,754 9
0,7	0,758 0	0,761 1	0,764 2	0,767 3	0,770 4	0,773 4	0,776 4	0,779 4	0,782 3	0,785 2
0,8	0,788 1	0,791 0	0,793 9	0,796 7	0,799 5	0,802 3	0,805 1	0,807 8	0,810 6	0,813 3
0,9	0,815 9	0,818 6	0,821 2	0,823 8	0,826 4	0,828 9	0,831 5	0,834 0	0,836 5	0,838 9
1,0	0,841 3	0,843 8	0,846 1	0,848 5	0,850 8	0,853 1	0,855 4	0,857 7	0,859 9	0,862 1
1,1	0,864 3	0,866 5	0,868 6	0,870 8	0,872 9	0,874 9	0,877 0	0,879 0	0,881 0	0,883 0
1,2	0,884 9	0,886 9	0,888 8	0,890 7	0,892 5	0,894 4	0,896 2	0,898 0	0,899 7	0,901 5
1,3	0,903 2	0,904 9	0,906 6	0,908 2	0,909 9	0,911 5	0,913 1	0,914 7	0,916 2	0,917 7
1,4	0,919 2	0,920 7	0,922 2	0,923 6	0,925 1	0,926 5	0,927 9	0,929 2	0,930 6	0,931 9
1,5	0,933 2	0,934 5	0,935 7	0,937 0	0,938 2	0,939 4	0,940 6	0,941 8	0,942 9	0,944 1
1,6	0,945 2	0,946 3	0,947 4	0,948 4	0,949 5	0,950 5	0,951 5	0,952 5	0,953 5	0,954 5
1,7	0,955 4	0,956 4	0,957 3	0,958 2	0,959 1	0,959 9	0,960 8	0,961 6	0,962 5	0,963 3
1,8	0,964 1	0,964 9	0,965 6	0,966 4	0,967 1	0,967 8	0,968 6	0,969 3	0,969 9	0,970 6
1,9	0,971 3	0,971 9	0,972 6	0,973 2	0,973 8	0,974 4	0,975 0	0,975 6	0,976 1	0,976 7
2,0	0,977 2	0,977 9	0,978 3	0,978 8	0,979 3	0,979 8	0,980 3	0,980 8	0,981 2	0,981 7
2,1	0,982 1	0,982 6	0,983 0	0,983 4	0,983 8	0,984 2	0,984 6	0,985 0	0,985 4	0,985 7
2,2	0,986 1	0,986 4	0,986 8	0,987 1	0,987 5	0,987 8	0,988 1	0,988 4	0,988 7	0,989 0
2,3	0,989 3	0,989 6	0,989 8	0,990 1	0,990 4	0,990 6	0,990 9	0,991 1	0,991 3	0,991 6
2,4	0,991 8	0,992 0	0,992 2	0,992 5	0,992 7	0,992 9	0,993 1	0,993 2	0,993 4	0,993 6
2,5	0,993 8	0,994 0	0,994 1	0,994 3	0,994 5	0,994 6	0,994 8	0,994 9	0,995 1	0,995 2
2,6	0,995 3	0,995 5	0,995 6	0,995 7	0,995 9	0,996 0	0,996 1	0,996 2	0,996 3	0,996 4
2,7	0,996 5	0,996 6	0,996 7	0,996 8	0,996 9	0,997 0	0,997 1	0,997 2	0,997 3	0,997 4
2,8	0,997 4	0,997 5	0,997 6	0,997 7	0,997 7	0,997 8	0,997 9	0,997 9	0,998 0	0,998 1
2,9	0,998 1	0,998 2	0,998 2	0,998 3	0,998 4	0,998 4	0,998 5	0,998 5	0,998 6	0,998 6

Table pour les grandes valeurs de u

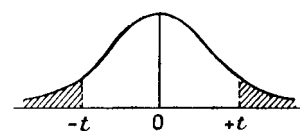
u	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
F(u)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,99984	0,999928	0,999968	0,999997

La table donne les valeurs de  $\Phi(u)$  pour u positif. Lorsque u est négatif il faut prendre le complément à l'unité de la valeur lue dans la table.

Exemple: pour  $u = + 1,21$ ,  $\Phi(u) = 0,886 9$ ; pour  $u = -1,21$ ,  $\Phi(u) = 0,113 1$

## DISTRIBUTION t

La table donne la probabilité  $\alpha$  pour que  $t$  égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).

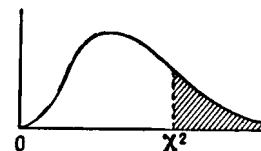


$\alpha$ d.d.l.	0,90	0,50	0,30	0,20	0,10	<b>0,05</b>	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
$\infty$	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Exemple: avec d.d.l. = 10, pour  $t = 2,228$  la probabilité est  $\alpha = 0,05$ .

## DISTRIBUTION $\chi^2$

La table donne la probabilité  $\alpha$  pour que  $\chi^2$  égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.)



$\alpha$									
d.d.l.	0,90	0,50	0,30	0,20	0,10	<b>0,05</b>	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	18,939	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	19,768	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	20,599	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,703

Exemple: avec d.d.l. = 3, pour  $\chi^2 = 7,815$  la probabilité est  $\alpha = 0,05$  (5%)