

Préparation à l'agrégation externe de Sciences Sociales

Statistique inférentielle - Tests d'hypothèses

2021-2022

Exercice 1

1. Le risque contrôlé de façon préférentielle est le risque de première espèce

$$\mathbf{P}_{H_0} \text{ (On rejette } H_0 \text{)}$$

qui correspond au fait d'innocenter à tort l'aéroport : les concepteurs du test jugent donc moins grave d'indemniser à tort les riverains que de commettre l'erreur inverse. Le risque de seconde espèce est justement la probabilité d'indemnisation à tort

$$\mathbf{P}_{H_1} \text{ (On ne rejette pas } H_0 \text{)}$$

2. (a) Sous H_0 , \overline{X}_{40} est la moyenne empirique de 40 variables indépendantes de même loi $\mathcal{N}(80, 49)$, donc suit la loi $\mathcal{N}(80, \frac{49}{40})$.
- (b) La région critique du test, aussi appelée *zone de rejet*, est l'ensemble des observations qui conduisent au rejet de l'hypothèse H_0 . On souhaite bien sûr rejeter H_0 si la moyenne empirique des observations réalisées est trop faible par rapport à 80, c'est-à-dire que la zone de rejet est de la forme

$$Z = \left\{ x \in \mathbb{R}^+ \mid \frac{1}{40} \sum_{i=1}^{40} x_i \leq t \right\}$$

pour un certain $t \in [0, 80]$. Sous H_0 , la probabilité pour que $(X_1, \dots, X_{40}) \in Z$ doit être égale à 1%, donc on doit avoir

$$\mathbf{P}_{H_0}(\overline{X}_{40} \leq t) = 0,01$$

soit

$$\mathbf{P}_{H_0} \left(\frac{\overline{X}_{40} - 80}{\sqrt{\frac{49}{40}}} \leq \frac{t - 80}{\sqrt{\frac{49}{40}}} \right) = 0,01$$

soit encore

$$\Phi \left(\frac{t - 80}{\sqrt{\frac{49}{40}}} \right) = 0,01$$

puisque $\frac{\overline{X}_{40} - 80}{\sqrt{\frac{49}{40}}} \sim \mathcal{N}(0, 1)$ d'après la question précédente, d'où

$$t = 80 + \sqrt{\frac{49}{40}} \Phi^{-1}(0,01) \approx 80 - 1,11 \times 2,33 \approx 77,41$$

grâce à la table 1, et enfin

$$Z = \left\{ x \in \mathbb{R}^+ \mid \frac{1}{40} \sum_{i=1}^{40} x_i \leq 80 + \sqrt{\frac{49}{40}} \Phi^{-1}(0,01) \right\}$$

FRACTILES DE LA LOI NORMALE RÉDUITE

P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,99
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537	0,88
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	0,87
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,86
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,85
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,84
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,83
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,82
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,81
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	0,80
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,89
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,88
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,87
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,86
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407	1,0364	0,85
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,84
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,83
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,82
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,81
0,19	0,8779	0,8742	0,8705	0,8669	0,8633	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,80
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,79
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,78
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,77
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,76
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,75
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,74
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,73
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,72
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,71
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,70
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4959	0,69
0,31	0,4959	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,88
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,67
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,66
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,65
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,64
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,63
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,62
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,61
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,60
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,59
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,58
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,57
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,56
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,55
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,54
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,53
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,52
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,51
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,50
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

Grandes valeurs de u

P	0,9999	0,99999	0,999999	0,9999999	0,99999999	0,999999999
u	3,7190	4,2649	4,7534	5,1993	5,6120	5,9978

FIGURE 1 – Table de quantiles de la loi normale centrée réduite

- On choisit de rejeter H_0 au profit de H_1 si les observations sont dans la région critique, c'est-à-dire si la valeur observée de X_{40} est plus basse que 77,41, et on conserve H_0 sinon.
- Il se trouve que $79 > 77,41$. On ne rejette donc pas l'hypothèse H_0 au niveau 1%.
- On prend ici les affirmations de la compagnie pour argent comptant. Les variables observées suivent donc, on le sait à présent, la loi $\mathcal{N}(78,49)$. Nous avons décidé d'indemniser les riverains dès lors que la moyenne des observations réalisées dépassait 77,41 ; la probabilité recherchée est donc la probabilité pour qu'une variable de loi $\mathcal{N}(78, \frac{49}{40})$ dépasse la valeur 77,41, c'est-à-dire

$$\mathbf{P}(\overline{X_{40}} > 77,41) = \mathbf{P}\left(\frac{\overline{X_{40}} - 78}{\sqrt{\frac{49}{40}}} > \frac{77,41 - 78}{\sqrt{\frac{49}{40}}}\right) \approx 1 - \Phi(-0,533) = \Phi(0,533) = 0,703$$

d'après la table donnée dans la figure 1.

Exercice 2

- Il suffit de multiplier l'effectif présent à la fin de chaque ligne par les fréquences données en colonnes en arrondissant à l'unité la plus proche.

$\hat{n}_{i,j}$	SR	Non	Oui (non précis)	Oui (précis)	Effectifs
Classes populaires	0	37	7	9	53
Classes moyennes	0	39	15	44	98
Classes supérieures	2	9	15	73	99

2. Il faut d'abord calculer les fréquences de chaque réponse sur la totalité de l'échantillon de 250 personnes, puis les multiplier par les fréquences d'appartenance à chaque classe sociale. Par exemple, la fréquence des non-réponses est égale à 0,008 et celle de l'appartenance aux classes populaires à 0,212, donc la fréquence théorique correspondante est $f_{1,1} = 0,008 \times 0,212 \approx 0,002$. On obtient le tableau suivant (dans lequel les fréquences ne se somment pas correctement à cause des erreurs d'arrondis) :

$f_{i,j}$	SR	Non	Oui (non précis)	Oui (précis)	f_{classe}
Classes populaires	0,002	0,072	0,031	0,107	0,212
Classes moyennes	0,003	0,133	0,058	0,198	0,392
Classes supérieures	0,003	0,135	0,059	0,200	0,396
$f_{\text{réponses}}$	0,008	0,34	0,148	0,504	

On obtient les effectifs théoriques correspondants en multipliant toutes les cases du tableau des fréquences théoriques par 250 :

$n_{i,j}$	SR	Non	Oui (non précis)	Oui (précis)
Classes populaires	0,5	18	7,75	26,75
Classes moyennes	0,75	33,25	14,5	49,5
Classes supérieures	0,75	33,75	14,75	50

3. La statistique du χ^2 est

$$C = \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{i,j} - \hat{n}_{i,j})^2}{n_{i,j}}$$

donc la contribution de chaque case (i, j) à cette statistique est égale à $\frac{(n_{i,j} - \hat{n}_{i,j})^2}{\hat{n}_{i,j}}$. Par exemple, la contribution de la case $(2, 2)$ (« Classes moyennes - Non ») est égale à

$$c_{2,2} = \frac{(39 - 33,25)^2}{33,25} \approx 0,994$$

Les contributions des différentes cases sont données ci-après :

$c_{i,j}$	SR	Non	Oui (non précis)	Oui (précis)
Classes populaires	0,5	20,056	0,726	11,778
Classes moyennes	0,75	0,994	0,017	0,611
Classes supérieures	2,083	18,15	0,004	10,58

4. L'hypothèse H_0 est l'hypothèse d'indépendance : si l'on note (X_k, Y_k) le couple de variables observées auprès de l'individu $k \in \{1, \dots, 250\}$, avec X_k représentant l'appartenance à une classe sociale et Y_k la réponse apportée à la question posée, alors

$$H_0 : \forall k \in \{1, \dots, 250\}, \quad X_k \text{ et } Y_k \text{ sont indépendantes.}$$

On choisit ensuite simplement $H_1 = \overline{H_0}$. Le risque de première espèce est alors la probabilité de rejeter à tort l'hypothèse d'indépendance.

On sait que sous H_0 , la loi suivie par la statistique C est proche de la loi $\chi_{(3-1)(4-1)}^2 = \chi_6^2$. On a donc pour tout $\alpha \in]0, 1]$ et $t \in \mathbb{R}_+$:

$$\mathbf{P}_{H_0} \left(C > F_{\chi_6^2}^{-1}(1 - \alpha) \right) = 1 - F_{\chi_6^2} \left(F_{\chi_6^2}^{-1}(1 - \alpha) \right) = \alpha$$

donc le test consistant à rejeter H_0 si et seulement si $C > F_{\chi_6^2}^{-1}(1 - \alpha)$ est un test de niveau α de H_0 contre H_1 . Notons que l'on choisit une zone de rejet du type $\{C > \dots\}$ puisque l'on veut rejeter l'hypothèse H_0 lorsque le contraste C est trop fort !

En choisissant par exemple $\alpha = 5\%$, on lit sur la figure 2 que l'on rejette H_0 si et seulement si $C > 12,59$. La question précédente permet de voir que sur les observations réalisées on a $C \approx 66,249$, et donc que ces observations permettent de rejeter H_0 .

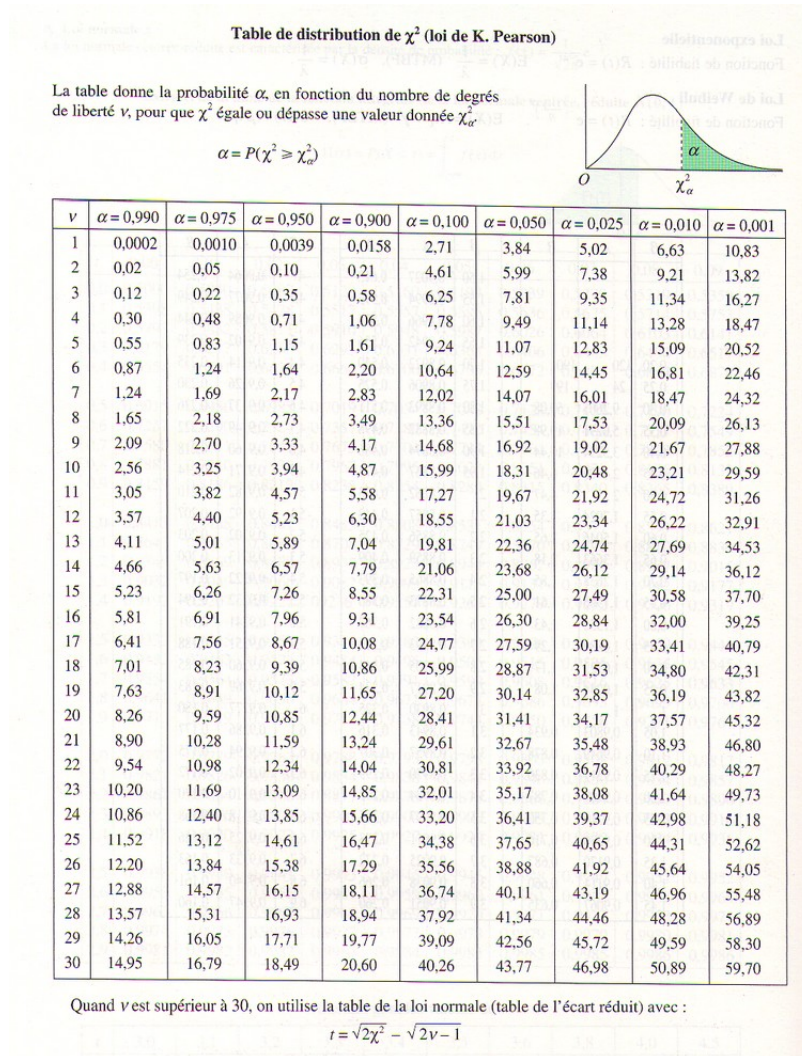


FIGURE 2 – Table de la loi du χ^2

- Pour calculer la p -valeur du test, il suffit de diminuer graduellement la valeur de α choisie ci-dessus et de relever à partir de quelle valeur de α il ne sera plus possible de rejeter le test au niveau α compte tenu des données dont on dispose. La figure 2 indique que cette valeur est (bien) inférieure à 0,001 puisque H_0 est encore rejeté au niveau 0,001. Cette observation conduit donc à rejeter très fortement (c'est-à-dire avec un très grand niveau de certitude) l'hypothèse d'indépendance. Pour calculer explicitement la p -value, il suffit de remarquer qu'elle est atteinte lorsque

$$66,249 = F_{\chi_6^2}^{-1}(1 - \alpha)$$

et donc qu'elle vaut $\alpha = 1 - F_{\chi_6^2}(66,249)$. Cette valeur est en réalité si proche de 0 qu'un logiciel de statistiques standard l'affiche avec une précision de 20 décimales comme étant égale à 0.

Insistons sur le fait que la p -value du test est dépendante des données observées et qu'elle ne peut en aucun cas être calculée indépendamment d'observations numériques concrètes !

Exercice 3

- Appliquons directement le résultat du cours relatif au test d'égalité de moyennes de lois gaussiennes de variances inconnues : si $\alpha \in]0, 1]$, on choisit de rejeter l'hypothèse $H_0 : \mu_1 = \mu_2$ en faveur de

l'hypothèse $H_1 = \overline{H_0}$ si et seulement si

$$\left| \frac{\overline{X_6} - \overline{Y_{300}}}{\sqrt{\frac{\sigma_X^2}{6} + \frac{\sigma_Y^2}{300}}} \right| > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right)$$

pour obtenir un test de niveau α de H_0 contre H_1 . Notons que ce résultat est rendu possible par le fait que les écarts-types théoriques des distributions des deux échantillons sont connus.

2. La p -value du test est la borne inférieure de l'ensemble des $\alpha \in]0, 1]$ tels que

$$\left| \frac{493 - 530}{\sqrt{\frac{187^2}{6} + \frac{300^2}{300}}} \right| > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right),$$

c'est-à-dire l'unique valeur α telle que

$$\left| \frac{493 - 530}{\sqrt{\frac{187^2}{6} + \frac{300^2}{300}}} \right| = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right),$$

soit

$$\Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \approx 0,473$$

ou encore

$$\alpha = 2(1 - \Phi(0,473)) \approx 0,6362.$$

Cette p -value est très haute : il est impossible de rejeter l'hypothèse d'égalité des moyennes avec un bon niveau de certitude.