Contrôle : Statistique inférentielle

Patrick Tardivel, Université de Bourgogne

27/06/2023

Exercice 1. *(6 points) On souhaite savoir si en moyenne le poids d'un individu a diminué après consultation d'un diététicien.*

Données : *Avant puis après la consultation chez le diététicien les poids des individus sont* $x_1^{\text{avant}}, x_2^{\text{avant}}, \dots, x_{220}^{\text{avant}}$ *et* $x_1^{\text{après}}, x_2^{\text{après}}, \dots, x_{220}^{\text{après}}$. *On a obtenu les résultats suivants :*

$$\sum_{i=1}^{220} x_i^{\text{avant}} = 13150,14\,, \sum_{i=1}^{220} x_i^{\text{après}} = 13035,50 \text{ et } \sum_{i=1}^{220} (x_i^{\text{avant}} - x_i^{\text{après}})^2 = 1775,38.$$

1. *Faire un test statistique au niveau* 5% *pour conclure en rédigeant soigneusement chaque étape.*
2. *Calculer la p-valeur.*
3. *Rappeler la démarche mathématiques permettant de construire un intervalle de confiance pour une probabilité de niveau asymptotique* $1-\alpha$.
4. *Sur les 220 patients, 133 ont réduit leurs poids après consultation du diététicien. Donner un intervalle de confiance de niveau asymptotique* 0,90 *pour la probabilité inconnue qu'un individu ait diminué sont poids après consultation du diététicien.*

Exercice 2. *(6 points)*

On désir savoir s'il existe une dépendance entre la note d'un établissement hôtelier et son emplacement en ville. Dans une ville, un échantillon d'hôtels évalués donne les résultats suivants.

	Très mauvais (0)	*Mauvais (1)*	*Correct (2)*	*Bien (3)*	*Excellent (4)*
En banlieue (0)	*7*	*17*	*36*	*35*	*19*
Au centre-ville (1)	*5*	*25*	*30*	*41*	*24*

1. *Faire un test statistique du khi-deux d'indépendance au niveau* 5% *pour conclure en rédigeant soigneusement chaque étape (le calcul de la p-valeur n'est pas demandé).*
2. *Est-ce que l'on observe une différence entre les proportions de bons hôtels (bien ou excellent) en banlieue et au centre-ville ?*
3. *Calculer la p-valeur.*

Exercice 3. *(8 points) On considère les variables aléatoires* A_0, A_1, B_0, B_1 *qui modélisent les longueurs d'un rectangle prenant en compte l'erreur de mesure. On note a la longueur du rectangle, b la largeur du rectangle et on suppose que* $\mathbb{E}(A_0) = \mathbb{E}(A_1) = a, \mathbb{E}(B_0) = \mathbb{E}(B_1) = b$ *et* $\text{var}(A_0) = \text{var}(A_1) = \text{var}(B_0) = \text{var}(B_1) = \sigma^2$.

Indications : les formules suivantes sont admises :

- *Si* X *et* Y *sont des variables aléatoires indépendantes alors,* $E(XY) = E(X)E(Y)$.
- *Si* $\mathbb{E}(X) = \mu$ *et* $\text{var}(X) = \sigma^2$ *alors* $\mathbb{E}(X^2) = \mu^2 + \sigma^2$.
- *Si* X *et* Y *sont des variables aléatoires indépendantes alors* X^2 *et* Y^2 *sont également des variables indépendantes.*
- *Si* W, X, Y *et* Z *sont des variables aléatoires indépendantes alors* $W+X$ *et* $Y+Z$ *sont également des variables indépendantes.*

1. *On considère l'estimateur* $T_1 = A_0 B_0$ *de ab.*
   (a) *Calculer le biais de* T_1 *pour ab.*
   (b) *Calculer l'erreur quadratique moyenne de* T_1 *pour ab.*
2. *On considère l'estimateur* $T_2 = \dfrac{(W+X)(Y+Z)}{4}$ *de ab.*

(a) *Calculer le biais de T_2 pour ab.*

(b) *Calculer l'erreur quadratique moyenne de T_2 pour ab.*

3. *Quel est le meilleur estimateur entre T_1 et T_2 ?*
4. *On considère l'estimateur $T_3 = \sqrt{A_0 A_1 B_0 B_2}$ de ab.*

   (a) *Calculer $\mathbb{E}(T_3^2)$.*

   (b) *Déduire de la question précédente que T_3 est un estimateur biaisé pour ab.*
5. *Pour cette question on suppose A_0, A_1 sont des variables aléatoires de loi $\mathcal{N}(10, 2^2)$ et que B_0, B_1 sont des variables aléatoires de loi $\mathcal{N}(20, 2^2)$. Écrire un programme qui à partir d'un échantillon expérimental de taille n (grand) des variables A_0, A_1, B_0, B_1 donne une approximation de $\mathbb{E}(T_3)$. Vous justifierez votre démarche.*

Formulaire

— Soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ un n-échantillon à valeur dans l'ensemble fini $D \times E$. Pour $i \in D$ et $j \in E$ on pose

$$N_{i*} = \sum_{k=1}^{n} \mathbf{1}(X_k = i) \quad N_{*j} = \sum_{k=1}^{n} \mathbf{1}(Y_k = j) \text{ et } N_{ij} = \sum_{k=1}^{n} \mathbf{1}((X_k, Y_k) = (i, j)).$$

Sous l'hypothèse nulle, lorsque X et Y sont indépendantes, on a la convergence en loi suivante :

$$W = \sum_{i \in D} \sum_{j \in E} \frac{\left(N_{ij} - \frac{N_{i*} N_{*j}}{n}\right)^2}{\frac{N_{i*} N_{*j}}{n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi((\text{card}(D) - 1)(\text{card}(E) - 1) \text{ ddl}).$$

— Soit $X_1, \ldots, X_n$ un n−échantillon d'une loi P inconnue à valeur dans un ensemble fini D et P^0 une loi de probabilité sur D. Pour $i \in D$ on pose $N_i = \sum_{j=1}^{n} \mathbf{1}(X_i = i)$. Sous l'hypothèse nulle, lorsque $P^0 = P$, on a la convergence en loi suivante :

$$Y = \sum_{i \in D} \frac{(N_i - nP^0(i))^2}{nP^0(i)} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi(\text{card}(D) - 1 \text{ ddl}).$$

— Soit $(D_1, \ldots, D_n)$ un n−échantillon d'une loi de moyenne $\mu \in \mathbb{R}$ et de variance σ^2. Sous l'hypothèse nulle, lorsque $\mu = 0$, on a la convergence en loi suivante :

$$Z = \frac{\overline{D}}{\sqrt{S^2_{corr}/n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

— Soit $(X_1, \ldots, X_n)$ un n−échantillon de loi $\mathcal{B}(p)$ alors :

$$\frac{\overline{X} - p}{\sqrt{\overline{X}(1 - \overline{X})/n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

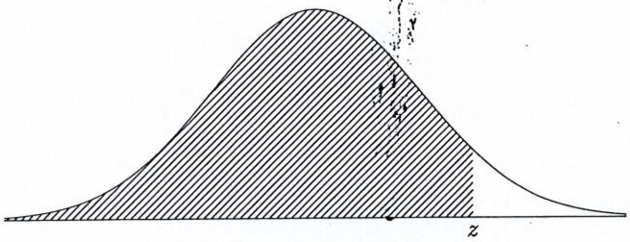
— Soit X une variable aléatoire de loi $\mathcal{B}(n_1, p_1)$ et Y une variable aléatoire de loi $\mathcal{B}(n_2, p_2)$. Sous l'hypothèse nulle, lorsque $p_1 = p_2$; on a la convergence en loi suivante :

$$Z = \frac{X/n_1 - Y/n_2}{\sqrt{\hat{p}_0(1 - \hat{p}_0)(1/n_1 + 1/n_2)}} \xrightarrow[\min\{n_1, n_2\} \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \text{ où } \hat{p}_0 = (X + Y)/(n_1 + n_2).$$

Table de la loi normale centrée réduite

Soit Z une variable aléatoire de loi $\mathcal{N}(0,1)$. Les valeurs de la fonction de répartition $z \in \mathbb{R} \mapsto \mathbb{P}(Z \leq z)$ sont données dans le tableau suivant :

densité d'une loi normale $\mathcal{N}(0,1)$



L'aire hachurée illustre $\mathbb{P}(Z \leq z)$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Pour $z \geq 3$ on peut utiliser l'approximation

$$\mathbb{P}(Z \leq z) \approx 1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\left(\frac{1}{z} - \frac{1}{z^3}\right).$$

Valeurs critiques au niveau 5% : loi normale centrée réduite

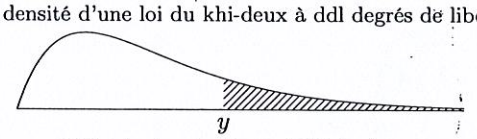
La valeur critique pour un test bilatéral de niveau 5% est $1,96$: $\mathbb{P}(|Z| \geq 1,96) = 0,05$.

La valeur critique pour un test unilatéral de niveau 5% est $1,645$: $\mathbb{P}(Z \geq 1,645) = 0,05$.

Valeurs critiques au niveau 5% loi du chi-carré

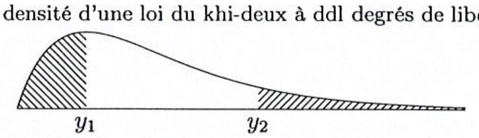
Soit Y une variable aléatoire de loi de chi-deux à ddl degrés de liberté. Les valeurs critiques pour effectuer un test unilatéral ou bilatéral au niveau 5% sont données au tableau suivant :

densité d'une loi du khi-deux à ddl degrés de liberté



L'aire hachurée illustre $\mathbb{P}(Y \geq y)$.

densité d'une loi du khi-deux à ddl degrés de liberté



L'aire hachurée illustre $\mathbb{P}(Y \leq y_1 \text{ ou } Y \geq y_2)$.

ddl	test unilatéral $\mathbb{P}(Y \geq y) = 0,05$	test bilatéral $\mathbb{P}(Y \leq y_1$ ou $Y \geq y_2) = 0,05$	
1	3,8414	0,0009	5,0238
2	5,9914	0,0506	7,3777
3	7,8147	0,2157	9,3484
4	9,4877	0,4844	11,1432
5	11,0704	0,8312	12,8325
6	12,5915	1,2373	14,4493
7	14,0671	1,6898	16,0127
8	15,5073	2,1797	17,5345
9	16,9189	2,7003	19,0227
10	18,3070	3,2469	20,4831
11	19,6751	3,8157	21,9200
12	21,0260	4,4037	23,3366