Introduction à la statistique non-paramétrique

DM – 2020 –

Exercice 1 : Nous nous intéressons à la hauteur des arbres d'une même espèce de chênes présente dans une forêt.

1. Nous supposons que la hauteur de ces chênes peut être modélisée par une variable aléatoire de loi normale $\mathcal{N}(\mu, \sigma^2)$. Nous disposons d'un échantillon de $n = 8$ mesures $x_1^n = (x_1, .., x_n)$ donnant une moyenne empirique $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i = 20.31$ mètres.
   (a) Nous supposons que σ est connu et égal à 5.
   Donner un intervalle de confiance de taille 95%, puis de taille 99%, pour μ.
   Proposer une procédure de test de $H_0 : \mu = 16$ contre $H_1 : \mu \neq 16$.
   Procéder au test au niveau $\alpha = 5\%$, puis au niveau $\alpha = 1\%$.
   (b) Nous supposons maintenant que σ est inconnu.
   L'écart-type empirique mesuré $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}$ est de 3.96.
   Proposer une procédure de test de $H_0 : \mu = 16$ contre $H_1 : \mu \neq 16$, puis de $H_0 : \mu \leq 16$ contre $H_1 : \mu > 16$.
   Déterminer dans chaque cas la p-valeur observée, et conclure si l'on teste au niveau $\alpha = 5\%$.

2. Nous disposons maintenant de la série complète des données x_1^n (hauteurs en mètres)

   19.37 21.12 24.84 18.10 15.61 24.05 14.17 25.26

   et nous cherchons à tester si cet échantillon est distribué selon une loi normale. Pour cela nous supposons que x_1^n est issu d'un échantillon i.i.d. X_1^n de fonction de répartition F et l'on note $\hat{F}_n$ la fonction de répartition empirique qui lui est associée. Nous notons N_{μ,σ^2} la fonction de répartition d'une loi normale de paramètres μ et σ^2. Nous cherchons à tester :

   $$H_0 : F \in \mathcal{F}N \quad \text{contre} \quad H_1 : F \notin \mathcal{F}N$$

   où $\mathcal{F}N = \{G \; ; \; \exists \, (\mu, \sigma^2) \in (\mathbb{R} \times \mathbb{R}_+^*) \;\text{ tel que }\; G = N_{\mu,\sigma^2}\}$.
   Nous rappelons qu'ici $\hat{\mu} = 20.31$ et $\hat{\sigma} = 3.96$, et nous utilisons comme statistique de test $\Delta_n = \sup_{t\in\mathbb{R}} |\hat{F}_n(t) - N_{\hat{\mu},\hat{\sigma}^2}(t)|$.
   (a) Rappeler la loi de cette statistique sous H_0 et la région de rejet du test.
   (b) Par une lecture graphique sur la Figure 1, où nous avons tracé $\hat{F}_n$ et $N_{\hat{\mu},\hat{\sigma}^2}$, déterminer la valeur de cette statistique.
   (c) La loi sous H_0 de cette statistique de test a été tabulée pour $n = 8$. En utilisant les quantiles donnés ci-dessous, conclure si l'on teste au niveau $\alpha = 5\%$.

	$q_{5\%}$	$q_{10\%}$	$q_{90\%}$	$q_{95\%}$
Statistique du test d'appartenance à la loi normale	0.14	0.15	0.27	0.29

   (d) Donner une méthode d'estimation des quantiles de cette loi.

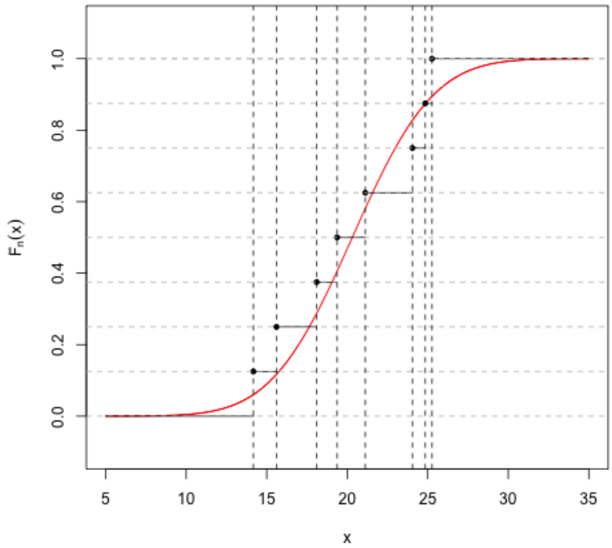

FIGURE 1 – $\hat{F}_n$ (noir) et $N_{\hat{\mu},\hat{\sigma}^2}$ (rouge)

Exercice 2 : On a mesuré le rythme cardiaque (en nombre de battements par minute), avant et après un don du sang, de 8 sujets pris au hasard parmi les personnes se présentant spontanément à un centre de transfusion sanguine. On cherche à savoir si le rythme cardiaque est plus faible après qu'avant le don du sang c'est-à-dire que l'on cherche à procéder au test H_0 : "le rythme cardiaque est le même avant et après le don du sang" contre H_1 : "le rythme cardiaque est plus faible après le don du sang qu'avant le don du sang".

1. Nous disposons tout d'abord uniquement de la donnée suivante : 6 personnes parmi les 8 avaient un rythme cardiaque plus faible après le don du sang qu'avant le don du sang.
   (a) Proposer une procédure de test
   (b) Donner la p-valeur observée. Conclure si l'on teste au niveau $\alpha = 5\%$.

2. Nous disposons à présent des données suivantes :

Sujet	1	2	3	4	5	6	7	8
Mesure avant le don du sang	78	101	77	75	92	76	66	87
Mesure après le don du sang	84	80	67	72	79	77	64	80

TABLE 1 – Nombre de battements cardiaques par minute

   (a) Proposer une procédure de test.
   (b) A l'aide de la Table 2, donner la p-valeur observée. Conclure si l'on teste au niveau $\alpha = 5\%$.

k	0	1	2	3	4	5	6	7	8	9
$P[\tilde{h}_8 \leq k]$	0.0039	0.0078	0.0117	0.0195	0.0273	0.0390	0.0547	0.0743	0.0978	0.1251
k	10	11	12	13	14	15	16	17	18	19
$P[\tilde{h}_8 \leq k]$	0.1563	0.1915	0.2306	0.2736	0.3204	0.3711	0.4220	0.4727	0.5274	0.5783
k	20	21	22	23	24	25	26	27	28	29
$P[\tilde{h}_8 \leq k]$	0.6290	0.6797	0.7265	0.7695	0.8087	0.8438	0.8751	0.9024	0.9259	0.9454
k	30	31	32	33	34	35	36			
$P[\tilde{h}_8 \leq k]$	0.9610	0.9728	0.9805	0.9883	0.9922	0.9961	1.0000			

TABLE 2 – Loi de $\tilde{h}_8 := \sum_{j=1}^{8} jY_j$, où $Y_1, .., Y_8 \overset{iid}{\sim} Be(\frac{1}{2})$