Université de Strasbourg
UFR de Mathématiques et d'Informatique
Han-Ping LI

Année 2019/2020
Statistique :
Etude de cas L3

Fiche 5 Tests d'hypothèses (suite)

Exercice 1 :

1-a) Générer une réalisation d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = 19, \sigma = 5$. Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée" au seuil de significativité de $\alpha = 0.10$.

1-b) Générer $M = 1000$ réalisations d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = 19, \sigma = 5$. Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée". au seuil de $\alpha = 0.10$.

Combien y-a-il des réalisations conduisent à rejeter $\mathbf{H}_0$?

2-a) Générer une réalisation d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi exponentielle $\mathcal{E}(\lambda)$ avec $rate = \lambda = 1$. Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée" au seuil de $\alpha = 0.10$.

2-b) Générer $M = 1000$ réalisations d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi exponentielle $\mathcal{E}(\lambda)$ avec $rate = \lambda = 1$.
Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée" au seuil de $\alpha = 0.10$.

Combien y-a-il des réalisations conduisent à conserver $\mathbf{H}_0$?

3) Générer $M = 1000$ réalisations d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = 0.5$.
Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée" au seuil de $\alpha = 0.10$.

Combien y-a-il des réalisations conduisent à conserver $\mathbf{H}_0$?

4) Générer $M = 1000$ réalisations d'un échantillon de taille $n = 25$ $(X_1, \ldots, X_{25})$ d'une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = 10$.

Effectuer un test de Shapiro :
$\mathbf{H}_0$: "l'échantillon est issu d'une population normalement distribuée" contre $\mathbf{H}_1$: "l'échantillon est issu d'une population non normalement distribuée" au seuil de $\alpha = 0.10$.

Combien y-a-il des réalisations conduisent à conserver $\mathbf{H}_0$?

4-c) Tracer le barplot de la loi de Poisson en question puis superposer le graphe de la densité d'une loi normale avec $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = \sigma^2 = 10$. Que constatez-vous ?

```
barplot( dpois(0:25, lambda=10), col="blue")...
```

Exercice 2 :

Le ministère de la santé souhaite évaluer l'impact de l'augmentation récente des prix des cigarettes, mise en place en vue de réduire la proportion de fumeurs réguliers dans la population p. On sait que la proportion de fumeurs réguliers s'élevait à $p_0 = 0.32$ avant la hausse des prix. On prélève (après la hausse des prix) un échantillon de 300 individus dans lequel 84 individus fument régulièrement des cigarettes.

1) Tester $\mathbf{H_0}$ $: p \leq 0.32$ contre $\mathbf{H_1}$ $: p > 0.32$ au seuil $\alpha = 0.05$.

2) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

3) Tester $\mathbf{H_0}$ $: p \geq 0.32$ contre $\mathbf{H_1}$ $: p < 0.32$ au seuil $\alpha = 0.05$.

4) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

Exercice 3 :

Un bureau d'étude déclare que 20% d'étudiants vont au cinéma chaque mois. Une amicale décide de réaliser une enquête basée sur un échantillon de taille $n = 150$ dont on note S le nombre d'étudiants qui vont au cinéma mensuellement. Elle constate que sur un échantillon de taille n =150, 21 personnes vont au cinéma chaque mois.

1) Quelle est la loi exacte de S dans un échantillon de taille n = 150 ? Quelle est la loi asymptotique lorsque n tend vers l'infini ?

2) Préciser les hypothèses nulle et alternative ainsi que l'interprétation des risques de première espèce et de second espèce.

3) Le résultat de l'enquête est-il conforme à l'hypothèse du bureau au seuil $\alpha = 0.10$?

4) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur

sa probabilité ?

Exercice 4 :

Un chercheur pense que la proportion de garçons à la naissance augmente dans des conditions économiques difficiles. Pour vérifier cette hypothèse, il a prélevé deux échantillons de taille n= 5000 deux périodes différentes. L'échantillon correspondant à une période de récession économique révèle que **52,56%** des nouveau-nés étaient des garçons alors que l'autre échantillon donne une proportion de **51,46 %**.

1) Effectuer un test au seuil $\alpha = 0.10$. Le test corrobore-t-il les convictions du chercheur ?

2) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

3) En conservant les proportions des naissances de garçons, déterminer la plus petite valeur de n telle que la conclusion est inverse au seuil $\alpha = 0.10$.

4) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

Exercice 5 :

Dans une étude sur la qualité de deux publicités, chacune a été diffusée dans une zone test spécifique 6 fois en une semaine. La semaine suivante, une enquête téléphonique a identifié les personnes qui ont vu les publicités. L'enquête a demandé à ces personnes d'énoncer le slogan de la publicité qu'ils avaient vue. Voici les résultats :

- 69 personnes se souviennent du slogan parmi 150 personnes ayant vu la publicité A;

- 70 personnes se souviennent du slogan parmi 200 personnes ayant vu la publicité B.

1) Tester l'hypothèse selon laquelle il n'y a pas d'écart entre les proportions de personnes se souvenant du slogan des publicités, au seuil de significativité de 0.05.

2) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

Exercice 6 :

Nous souhaitons comparer la taille du lobe frontal des "crabes bleus" et celle des "crabes oranges". Pour simplifier l'étude, notre population des crabes bleus est constituée des 100 valeurs (en mm) de la variable "la taille du lobe frontal", notée "FL" du fichier "**crabs**" de la librairie "MASS" dont l'espèce (la couleur), notée "sp", est **B** et notre population des crabes orange est constituée des 100 valeurs (en mm) de la même variable "FL" des crabes dont l'espèce (la couleur), notée "sp", est **O** du même fichier.

Sauvegarder les 100 valeurs de "FL" des crabes bleus ("**B**") dans un vecteur nommé **Population1** et les 100 valeurs de "FL" des crabes orange ("**O**") dans un vecteur nommé **Population2**. .

Nous voulons savoir si la taille de crabes bleus est significativement plus petite que celle des crabes "orange" ou elles sont compatibles.

Pour ce faire, nous prélevons un échantillon de taille $n_1 = 28$ de la première population, et un échantillon de taille $n_2 = 29$ de la seconde population. Nous notons μ_1 et μ_2 les deux espérances de la variable "FL" selon la couleur des crabes.

Sauvegarder alors le vecteur $x = (x_1, \ldots, x_{28}))$ qui est une réalisation d'un échantillon de la première population ("crabes bleus") et le vecteur $y = (y_1, \ldots, y_{29})$ qui est une réalisation d'un échantillon de la deuxième population ("crabes oranges").

1. Effectuer trois tests préliminaires au seuil de $\tilde{\alpha} = 0,10$. Quelles seront vos conclusions ?

   On suppose dans la suite que les deux populations admettent pour lois deux lois normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$ respectivement avec les quatre paramètres inconnus.

2. Formuler l'hypothèse nulle $\mathbf{H}_0$ ainsi que l'hypothèse opposée $\mathbf{H}_1$.

3. Effectuer le test d'hypothèse principal, sans utiliser la commande "t.test()" en R, avec le seuil $\alpha = 0,05$.

   - Préciser la forme de la statistique utilisée.

   - Déterminer ses degrés de liberté.

4. - Déterminer la (ou les) valeur(s) critique(s).

5. - Conclure.

Exercice 7

Une agence environnementale soupçonne que le poisson présent dans un cours d'eau proche d'une usine a une concentration de mercure élevée. Pour confirmer cette suspicion, 37 poissons ont été capturés dans ce cours d'eau et la concentration en mercure de leur tissu a été mesurée. Un autre échantillon de 43 poissons d'un autre cours d'eau loin de toute usine ont également été capturés et et la concentration en mercure de leur tissu a été mesurée. Les concentrations en mercure des tissus du poisson en mg/kg sont indiquées ci-dessous (vous pouvez télécharger les données sur le site Moodle "Etude de Cas") :

proche de l'usine

0.19	0.63	0.92	0.80	0.47	1.19	0.66	0.88	0.59	0.08	0.98	1.01	0.45
0.93	1.06	0.79	0.50	0.71	0.89	0.61	0.91	0.76	1.30	0.15	1.02	0.59
0.79	0.90	0.69	0.50	0.13	0.11	0.80	0.82	1.01	1.09	0.68		

loin de toute usine

0.24	0.65	0.51	0.66	0.48	0.17	0.55	0.31	0.05	1.16	0.98	0.73	0.02
0.45	0.07	0.34	0.39	0.34	0.54	0.33	1.04	0.09	0.47	0.77	0.47	0.97
0.17	0.10	0.15	0.96	0.65	0.30	0.23	0.58	0.43	0.51	0.45	0.22	0.42
0.66	0.71	0.28	0.10									

On souhaite savoir si effectivement la concentration de mercure est plus élevée parmi les poissons vivant dans un cours d'eau proche d'une usine.

1) Effectuer les trois tests préliminaires

2) Effectuer un test au seuil $\alpha = 0.05$. Le test corrobore-t-il la crainte de l'agence ?

3) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

Exercice 8 :

Pour étudier l'évolution des heures passées à regarder la télévision dans un ménage, un institut de sondage a interrogé 250 ménages, a obtenu une moyenne de 7,6 heures par jour avec un écart-type de 2,6 heures. Un an plus tôt, les résultats donnaient une moyenne de 7,1 heures par jour avec un écart-type de 2.15 heures sur un échantillon de 200 ménages.

1) Que peut on conclure au seuil de $\alpha = 0.05$?

4) Quelle espèce d'erreur risque-t-on de commettre ? Que peut-on dire sur sa probabilité ?

Exercice 9 :

Soit $(X_1, \ldots, X_{20})$ un échantillon de taille $n = 20$ d'une loi normale $\mathcal{N}(\mu, \sigma^2)$. On considère le test bilatéral suivant : $\mathbf{H_0}$: $\sigma^2 = 4$ contre $\mathbf{H_1}$: $\sigma^2 \neq 4$ au seuil $\alpha = 0,05$.

- Déterminer la valeur critique ainsi que la zone de rejet du test.
- Que vaut la valeur de référence σ_0^2 dans l'expression suivante :

$$K^2 = \frac{(n-1)S_c^2)}{\sigma_0^2}?$$

On se propose d'étudier **le risque de deuxième espèce**. Les données doivent être choisies de sorte que $\mathbf{H_0}$ soit fausse.

I. Dans un premier temps, on simule M=10000 réalisations d'un échantillon avec $\mu = 10$ et $\sigma = 3$. On n'est pas censé de connaitre les deux paramètres $\mu = 10$ et $\sigma = 3$. On sauvegarde les 10000 réalisations des statistiques $\overline{X}$, S_c ainsi que celles de $K^2 = \frac{(n-1)S_c^2}{4}$..

1. A partir des 10000 réalisations de K^2, évaluer le risque. De quel risque s'agit-il ?
2. Comparer avec la valeur théorique donnée par (β_1=round(pchisq(4*32.85233/3 , df = 19) - pchisq(4*8.906516/3 , df = 19), 3))

II. Dans un deuxième temps, on simule M=10000 réalisations d'un échantillon avec $\mu = 10$ et $\sigma = 6$. On n'est pas censé de connaitre les deux paramètres $\mu = 10$ et $\sigma^= 6$. On sauvegarde les 10000 réalisations des statistiques $\overline{X}$, S_c ainsi que celles de $K^2 = \frac{(n-1)S_c^2}{4}$..

1. A partir des 10000 réalisations de K^2, évaluer le risque. De quel risque s'agit-il ? Commenter vos résultats.
2. Comparer avec la valeur théorique donnée par (β_2=round(pchisq(4*32.85233/6 , df = 19) - pchisq(4*8.906516/6 , df = 19), 3))