L2 - Psychologie 2019-2020 - Deux exercices sur le test binomial

Octobre 2019

Exercice 1. Les résultats au test de "Comment réagissez vous sous la pression" de Carter et Russell (voir le livre de Carter et Russell "More Psychometric Tests, 2003") montrent qu'une proportion de 45% d'adultes ont un résultat supérieur à 50. On fait passer le même test à 25 cadres d'une grande entreprise. Lors de ce test, 13 d'entre eux ont un résultat supérieur à 50.

1. Peut-on en déduire, au niveau de risque $\alpha = 5\%$, que les cadres de cette entreprise réagissent mieux sous la pression que la population générale ?
2. Déterminer la p-value du test.
3. Calculer la puissance $\eta\left(\frac{14}{25}\right)$.

Exercice 2. On cherche à déterminer si les enfants en bas âge ont une préférence pour une couleur. Pour cela, on fait une expérience avec un groupe d'enfants de 2 ans. Chaque enfant se voit proposer deux seaux, l'un rouge et l'autre vert et se doit de choisir l'un des deux.

1. La première expérience donne le résultat suivant : sur les 19 enfants de l'échantillon, 15 ont choisi le seau rouge. Peut-on conclure avec un risque inférieur à 5% que les enfants en bas âge préfèrent la couleur rouge à la couleur verte ?
2. Quelle est la p-value de ce test ?
3. Calculer la puissance $\eta\left(\frac{15}{19}\right)$.

Corrigé de l'Exercice 1. Les résultats au test de "Comment réagissez vous sous la pression" de Carter et Russell (voir le livre de Carter et Russell "More Psychometric Tests, 2003") montrent qu'une proportion de 45% d'adultes ont un résultat supérieur à 50. On fait passer le même test à 25 cadres d'une grande entreprise. Lors de ce test, 13 d'entre eux ont un résultat supérieur à 50.

1. Peut-on en déduire, au niveau de risque $\alpha = 5\%$, que les cadres de cette entreprise réagissent mieux sous la pression que la population générale ?

a) Hypothèses du test.

H_0 : les cadres de cette entreprise réagissent sous la pression comme la population générale

H_1 : les cadres de cette entreprise réagissent mieux sous la pression que la population générale

On désigne par p la proportion de cadres de l'entreprise qui ont un résultat supérieur à 50 au test de Carter et Russell.

$$H_0 : p = p_0 (= 0.45)$$
$$H_1 : p > p_0.$$

b) Statistique du test. On a $n = 25$, on est en présence d'un *petit échantillon*. On sait alors que, sous l'hypothèse nulle H_0, le nombre *aléatoire* S_n de cadres qui ont une note supérieure à 50 au test de Carter et Russell vérifie :

$$S_n \hookrightarrow \mathcal{B}(25, 0.45).$$

c) Région critique au niveau d'erreur $\alpha = 0.05$.

i) - TEXAS

(a) si, sur son modèle de calculatrice, les numéros de ligne n'apparaissent pas à gauche des colonnes, on conseille remplir la première colonne des numéros de 1 à 25 par l'ordre :seq(X,X,1,25,1). Cette colonne jouera le rôle de numéros de lignes.

(b) remplir la colonne suivante des nombres de 0 à 24, en mettant son curseur tout en haut de la deuxième colonne, et la commande $L1 - 1$ (OU BIEN, si on n'a pas eu besoin de faire l'étape a), on remplir la première colonne des nombres de 0 à 24, avec la commande 2ND→List→OPS→seq(X,X,0,24,1)).

(c) Dans la colonne suivante : Dist→binomcdf(25,.45,L_2) (OU BIEN Dist→binomcdf(25,.45,L_1) si on n'a pas eu besoin de faire l'étape a)). Cette colonne contient maintenant les nombres $\mathbb{P}[S_n < k]$. Par exemple, à la ligne 10, on a le nombre $\mathbb{P}[X < 10] = 0.242$.

(d) Dans colonne suivante : $1 - L_3$ (OU BIEN $1 - L_2$ si on n'a pas eu besoin de faire l'étape a)). Cette colonne contient maintenant les nombres $\mathbb{P}[S_n \geq k]$. Par exemple, la ligne 12 contient $\mathbb{P}[S_n \geq 12] = 0.457$.

On voit que $\mathbb{P}[S_n \geq 16] = 0.044 < 0.05$ et $\mathbb{P}[S_n \geq 15] = 0.096 > 0.05$, donc la région critique est $K_\alpha = \{S_n \geq 16\}$.

ii) CASIO.

(a) Dans le menu STAT, on remplit la première colonne des nombres de 0 à 23 par OPTN→LIST→Seq(X,X,0,24,1).

(b) On remplit la deuxième colonne des nombres $\mathbb{P}[X < k]$ par la commande Dist→Binom→Bcd→Choisir List (et pas Var), List L2→Numtrial:25→P:0.45→Save Res:List 2→Execute→Exit→Exit. Par exemple, à la ligne 10, on a le nombre $\mathbb{P}[S_n < 10] = 0.242$.

(c) On remplit la colonne 3 des nombres $\mathbb{P}[X \geq k]$ par la commande : 1-List 2. Par exemple, la ligne 12 contient le nombre $\mathbb{P}[S_n \geq 12] = 0.457$.

On voit que $\mathbb{P}[S_n \geq 16] = 0.044 < 0.05$ et $\mathbb{P}[S_n \geq 15] = 0.096 > 0.05$, donc la région critique est $K_\alpha = \{S_n \geq 16\}$.

REMARQUE. Sur certains anciens modèles de CASIO, on peut afficher les résultats $\mathbb{P}[X < k]$ de l'étape (b) mais pas les stocker dans une liste, et donc encore moins faire une troisième colonne qui contiendra les nombres $\mathbb{P}[X \geq k]$. Pour déterminer la région critique, on repère donc la ligne où les résultats basculent de "plus petits que 0.95 à "plus grands que 0.95". On observe alors, et on rédige de la façon suivante :

$$\mathbb{P}[X < 15] = 0.904 < 0.95, \text{ donc } \mathbb{P}[X \geq 15] > 0.05$$

et

$$\mathbb{P}[X < 16] = 0.956 > 0.95, \text{ donc } \mathbb{P}[X \geq 16] < 0.05.$$

On retrouve bien que la région critique est $K_\alpha = \{S_n \geq 16\}$.

d) Décision du test. Le résultat expérimental $n_e = 13$ n'appartient pas à la région critique K_α. Donc, au niveau d'erreur $\alpha = 0.05\%$, on conserve l'hypothèse H_0.

2. Déterminer la p-value du test. La p-value est la probabilité de la région de valeurs favorables à H_1 (c'est à dire les grandes valeurs de S_n) qui est bordée par la valeur expérimentale. Donc p-value$=\mathbb{P}[S_n \geq 13] = 0.306$.

Sur les modèles standard de TEXAS et CASIO, ce résultat est affiché à la ligne 13 de la dernière colonne remplie.

Sur les modèles anciens, on utilise les résultats des nombres $\mathbb{P}[S_n < k]$ qu'on vient de calculer :

$$p\text{-value} = \mathbb{P}[S_n \geq 13] = 1 - \mathbb{P}[S_n < 13] = 1 - 0.6936 = 0.306.$$

Le fait que la p-value soit supérieure à 0.05 confirme le fait que l'on doive conserver l'hypothèse H_0. Noter qu'il est bon de faire les calculs intermédiaires avec des nombres à 4 décimales, pour fournir le résultat final en arrondissant à 3 décimales.

3. Calculer la puissance $\eta\left(\frac{14}{25}\right)$. Il s'agit de calculer la probabilité de la région critique K_α *avec une loi de probabilité différente*, qui est $\mathcal{B}\left(25, \frac{14}{25}\right) = \mathcal{B}(25, 0.560)$. On trouve :

$$\eta\left(\frac{14}{25}\right) = \mathbb{P}[S_n \geq 16; \mathcal{B}(25, 0.560)] = 1 - \mathbb{P}[S_n \leq 15; \mathcal{B}(25, 0.560)] = 0.275.$$

Interprétation : Si la *proportion réelle* de cadres de l'entreprise qui ont une bonne réaction sous la pression est effectivement de $\frac{14}{25} \sim 0.560 = 56\%$, alors notre test a 0.2750=27.50% de chances d'accepter (*avec raison !*) l'hypothèse H_1.

Corrigé de l'Exercice 2. On cherche à déterminer si les enfants en bas âge ont une préférence pour une couleur. Pour cela, on fait une expérience avec un groupe d'enfants de 2 ans. Chaque enfant se voit proposer deux seaux, l'un rouge et l'autre vert et se doit de choisir l'un des deux.

1. La première expérience donne le résultat suivant : sur les 19 enfants de l'échantillon, 15 ont choisi le seau rouge. Peut-on conclure avec un risque inférieur à 5% que les enfants de bas âge préfèrent la couleur rouge à la couleur verte ? On procède à un test binomial.

a) Hypothèses du test.

H_0 : les enfants en bas-âge n'ont pas de préférence particulière entre les couleur la couleurs rouge et verte.

H_1 : les enfants en bas-âge préfèrent la couleur rouge.

On note p la proportion d'enfants en bas-âge qui préfèrent la couleur rouge.

$H_0 : p = p_0 (= 0.5)$; $H_1 : p > p_0$.

b) Statistiques du test. $n = 19$, il s'agit donc d'un petit échantillon. On sait alors que le nombre aléatoire S_n d'enfants qui préfèrent la couleur rouge sur un échantillon aléatoire de taille n vérifie:

$$S_n \hookrightarrow \mathcal{B}(19, 0.5).$$

c) Région critique au niveau $\alpha = 0.05$. A l'aide des méthodes détaillées dans le corrigé de l'Exercice 1, on trouve que

$$\mathbb{P}[S_n \geq 14] = 0.032 \text{ et } \mathbb{P}[S_n \geq 13] = 0.08,$$

donc $K_\alpha = \{S_n \geq 14\}$.

d) Le résultat expérimental $n_e = 15$ appartient à K_α, donc, au niveau d'erreur $\alpha = 0.05$, on accepte l'hypothèse H_1 : les enfants en bas-age préfèrent la couleur rouge à la couleur verte.

2. Quelle est la p-value de ce test ? C'est la probabilité de la région de valeurs favorables à H_1 (c'est à dire les grandes valeurs de S_n) et bordée par la valeur expérimentale $n_e = 15$. Donc $p - \text{value} = \mathbb{P}[S_n \geq 15] = 0.0096 \sim 0.010$. C'est une petite valeur, donc le test est significatif.

3. Calculer la puissance $\eta\left(\frac{15}{19}\right)$. C'est la probabilité d'acceptation de H_1 si la *proportion réelle* d'enfants préférant la couleur rouge est $\frac{15}{19} = 0.789 > 0.5$ (et doncc H_1 est satisfaite). C'est donc la propabilité d'appartenance à la région critique K_α pour la loi binomiale $\mathcal{B}(19, 0.7895)$. On a donc :

$$\eta\left(\frac{15}{19}\right) = \mathbb{P}\left[S_n \geq 14; \mathcal{B}\left(19, \frac{15}{19}\right)\right] = \mathbb{P}[S_n \geq 14; \mathcal{B}(19, 0.7895)] = 1 - \mathbb{P}[S_n \leq 13; \mathcal{B}(19, 0.7895)] = 0.8055.$$

Interprétation. Si la *proportion réelle* d'enfants en bas-âge qui préfèrent la couleur rouge est de $\frac{15}{19} \sim 0.7895 = 78.96\%$, alors notre test binomial a 80.55% de chances d'accepter (*avec raison !*) l'hypothèse H_1.