CONFORMITÉ D'UNE PROPORTION, QUELLE DÉMARCHE?

G.R.E.S.

Le « test de conformité d'une proportion », figure dans le programme du module D11 (Mathématiques appliquées et statistiques) en filière BTS de l'enseignement agricole. Sa mise en œuvre avec les étudiants n'est pas toujours sans embûche dès qu'il s'agit de déterminer la loi de probabilité à partir de laquelle on élabore la règle de décision.

L'objectif de l'article est de présenter trois possibilités d'élaboration de cette règle : à partir de la loi de probabilité « exacte » de la variable de décision retenue ou d'une loi normale qui permet d'approcher les valeurs numériques de la loi « exacte », l'approximation étant réalisée avec puis sans correction de continuité.

Exercice

Une machine est considérée opérationnelle si sa fabrication ne contient pas plus de 6% de pièces défectueuses.

On prélève au hasard dans cette fabrication un échantillon de 200 pièces. La fabrication étant supposée de grande taille, on admet que cet échantillon est un échantillon aléatoire simple. Il contient 17 pièces défectueuses.

Est-il possible, au seuil de risque 0,05, de conclure que la machine n'est pas opérationnelle ?

Notations :

Soit p la proportion de pièces défectueuses de la fabrication. Ce nombre p est une *valeur certaine* mais *inconnue*.

Soit F la variable aléatoire qui, pour chaque échantillon aléatoire simple de taille 200, extrait de la fabrication, prend pour valeur la *proportion* de pièces de cet échantillon qui sont défectueuses.

Première proposition de corrigé

Nous envisageons dans cette proposition de prendre en compte la loi de probabilité « exacte » de la variable de décision retenue en mettant en œuvre un **test binomial**, qui n'est *pas explicitement au programme* du module D11.

1. ***Formulation des hypothèses du test:***

$\mathbf{H_1}$: "la machine n'est pas opérationnelle" (hypothèse déduite de la question posée),

$\mathbf{H_0}$: "la machine est opérationnelle".

C'est-à-dire $H_1 : p > 0{,}06$;

$H_0 : p = 0{,}06$.

2. ***Nature du test*** :

D'après la formulation de H_1 le test est **unilatéral** à droite.

3. ***Variable de décision sous*** H_0 :

La variable aléatoire nF, que nous notons X, est la variable aléatoire qui, pour chaque échantillon aléatoire simple de taille 200, extrait de la fabrication, prend pour valeur le *nombre* de pièces de cet échantillon qui sont défectueuses. Cet échantillon étant un échantillon aléatoire simple, X est de loi *binomiale*.

Sous **l'hypothèse** $\mathbf{H_0}$, la variable aléatoire X est de loi binomiale $B(n\ ;\ p_0)$ avec $n = 200$ et $p_0 = 0{,}06$.

→ Choisissons X pour variable de décision.

Cet article est extrait du bulletin n°10 du Groupe de Réflexion sur l'Enseignement de la Statistique dans l'enseignement agricole (GRES).
Adresse du site : http://enfa.mip.educagri.fr/enfadraf/gres/.

ENFA : Ecole Nationale de Formation Agronomique.
G.R.E.S. Groupe de Réflexion sur l'Enseignement des Statistiques. (dans l'enseignement agricole)

4. ***Valeur critique*** :
Le seuil de signification du test est $\alpha = 0{,}05$. Le test est unilatéral à droite. Donc la valeur critique pour X est le plus petit nombre entier naturel k tel que $P(X \geqslant k) \leqslant 0{,}05$.
Ce nombre, que nous notons k_c, est donc le plus petit nombre entier naturel k tel que $P(X < k) \geqslant 0{,}95$ c'est-à-dire tel que $P(X \leqslant k - 1) \geqslant 0{,}95$ car X est une variable aléatoire *discrète*.

• ***Usage du tableur Excel pour déterminer*** k_c :
Détermination par balayage :

	1	**2**	**3**
1		n	200
2		p	0,06
3			
4		$k-1$	$P(X \leqslant k-1)$
5		17	0,9429
6		18	0,9672

LOI.BINOMIALE(LC(-1);L1C3;L2C3;1)

Détermination directe :

	1	**2**	**3**
1		n	200
2		p	0,06
3		$1-\alpha$	0,95
4			
5		$k_c - 1$	18

CRITERE.LOI.BINOMIALE(L1C3;L2C3;L3C3)

On obtient donc $k_c = 19$. C'est la valeur critique pour la variable aléatoire X.

5. ***Règle de décision*** :
Relativement à la variable X, la zone de rejet de l'hypothèse H_0 au seuil de risque 0,05 est [19 ; 200].

6. ***Décision*** :
L'échantillon prélevé contient 17 pièces défectueuses. Ce nombre n'appartient pas à la zone de rejet de H_0 . Donc, au seuil de risque 0,05, au vu des 200 observations, il n'y a pas lieu de rejeter H_0.

> Au seuil de risque 0,05, au vu des 200 observations, on ne peut pas conclure que la machine n'est pas opérationnelle.

Deuxième proposition de corrigé

Dans la démarche précédente, nous avons utilisé la loi de probabilité « exacte » de la variable de décision retenue. La taille de l'échantillon prélevé est assez grande ($n = 200$) pour envisager une approximation de cette loi par une loi normale.
La mise en œuvre « usuelle » d'un test de conformité d'une proportion consiste à réaliser une approximation sans correction de continuité de la loi de probabilité « exacte » de la variable aléatoire discrète F par une loi normale. On peut présenter cette mise en œuvre de la façon suivante :

Mêmes étapes **1.** et **2.** que dans la première proposition de corrigé.

3. ***Variable de décision sous*** $\mathbf{H_0}$:
Sous **l'hypothèse $\mathbf{H_0}$** , nF est de loi binomiale $B(n\ ;\ p_0)$ avec $n = 200$ et $p_0 = 0{,}06$.
L'espérance et la variance de nF sont respectivement :
$E(nF) = np_0$ et $V(nF) = np_0(1 - p_0)$.
Donc celles de F sont respectivement :
$E(F) = p_0 = 0{,}06$

et $V(F) = \dfrac{p_0(1-p_0)}{n} = 0{,}000282$.

→ Choisissons F pour variable de décision.

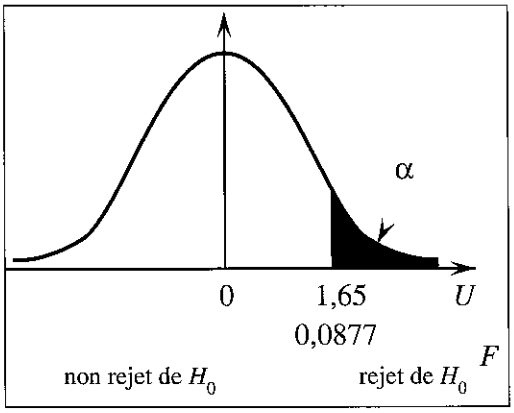
***n* est assez grand** pour admettre que la loi de probabilité de la variable aléatoire F peut être approchée par une loi normale.

Sous **l'hypothèse $\mathbf{H_0}$**, cette loi est la loi normale $N(0{,}06\ ;\ \sqrt{0{,}000282}\ )$.

Il en résulte que la variable aléatoire U, définie par $U = \dfrac{F - 0{,}06}{\sqrt{0{,}000\,282}}$, est approximativement de loi normale $N(0 ; 1)$.

4. *Schéma de décision et valeur critique* :

Il s'agit de représenter graphiquement le seuil de signification α du test et la zone de rejet de H_0 et de déterminer la valeur critique pour U.



Seuil de signification : $\alpha = 0{,}05$.
Le test est unilatéral à droite.
D'après la table de la fonction de répartition de U :

$$P(U > 1{,}64) > 0{,}05$$
$$P(U > 1{,}65) < 0{,}05$$

La valeur critique pour U est donc 1,65.
D'où la valeur critique pour F :

$$0{,}06 + 1{,}65 \times \sqrt{0{,}000\,282} = 0{,}0877$$

à 10^{-4} près

5. *Règle de décision* :

Relativement à la variable F, la zone de rejet de l'hypothèse H_0 au seuil de risque 0,05 est $[0{,}0877 ; 1]$.

6. *Décision* :

La proportion de pièces défectueuses dans l'échantillon prélevé est égale à 17/200 soit 0,085.
Cette proportion n'appartient pas à la zone de rejet de H_0.
Donc, au seuil de risque 0,05, au vu des 200 observations, il n'y a pas lieu de rejeter H_0.

Au seuil de risque 0,05, au vu des 200 observations, on ne peut pas conclure que la machine n'est pas opérationnelle.

Troisième proposition de corrigé

L'application de la **correction de continuité** lors de l'approximation de la loi de probabilité de F par une loi normale, permet "d'améliorer" cette approximation. On peut alors présenter la mise en œuvre du test de la façon suivante :
Mêmes étapes **1.** et **2.** que dans la première proposition de corrigé.

3. *Variable de décision sous* $\mathbf{H_0}$:

Sous **l'hypothèse $\mathbf{H_0}$**, la variable aléatoire nF, que nous notons X, est de loi binomiale $B(n ; p_0)$ avec $n = 200$ et $p_0 = 0{,}06$. L'espérance mathématique et la variance de X sont respectivement :
$E(X) = np_0 = 12$

et $V(X) = np_0(1 - p_0) = \sqrt{11{,}28}$.

→ Choisissons X pour variable de décision.
***n* est assez grand** pour admettre que la loi de probabilité de la variable aléatoire X peut être approchée par une loi normale. Sous **l'hypothèse $\mathbf{H_0}$**, cette loi est la loi normale $N(12 ; \sqrt{11{,}28})$.
Soit Z une variable aléatoire de loi normale $N(12 ; \sqrt{11{,}28})$. Il en résulte que la variable aléatoire U définie par

$U = \dfrac{Z - 12}{\sqrt{11{,}28}}$ est de loi normale centrée réduite $N(0 ; 1)$.

4. *Valeur critique pour* X :

Le seuil de signification du test est $\alpha = 0{,}05$. Le test est unilatéral à droite. Donc, la valeur critique k_c pour la variable aléatoire X est le plus petit nombre entier naturel non nul k, tel que

$P(X \geqslant k) \leqslant 0{,}05$.

En appliquant la correction de continuité lors de l'approximation de la loi de probabilité de X, on a :

$$P(X \geqslant k) \approx P(Z > k - 0{,}5).$$

Donc la valeur critique k_c est le plus petit nombre entier naturel non nul k tel que $P(Z > k - 0{,}5) \leqslant 0{,}05$ c'est-à-dire tel que $P(U > \dfrac{k - 0{,}5 - 12}{\sqrt{11{,}28}}) \leqslant 0{,}05$.

La fonction de répartition de U est **strictement croissante**, donc k_c est le plus petit nombre entier naturel non nul k tel que $\dfrac{k - 0{,}5 - 12}{\sqrt{11{,}28}} > 1{,}65$, soit $k > 18{,}04$. D'où $k_c = 19$. La valeur critique pour X est 19.

5. *Règle de décision* :

Relativement à la variable X, la zone de rejet de l'hypothèse H_0 au seuil de risque 0,05 est [19 ; 200].

6. *Décision* :

L'échantillon prélevé contient 17 pièces défectueuses. Ce nombre n'appartient pas à la zone de rejet de H_0. Donc, au seuil de risque 0,05, au vu des 200 observations, il n'y a pas lieu de rejeter H_0.

Au seuil de risque 0,05, au vu des 200 observations, on ne peut pas conclure que la machine n'est pas opérationnelle.

Conclusion

Pour comparer ces trois démarches, on peut, par exemple, exprimer la zone de rejet de l'hypothèse relativement à la variable aléatoire binomiale nF :

Démarche	Zone de rejet de H_0
1) Approximation (numérique) de la loi de nF sans correction de continuité.	[18 ; 200]
2) Approximation (numérique) de la loi de nF avec correction de continuité.	[19 ; 200]
3) Pas d'approximation de la loi de nF.	[19 ; 200]

Par contre, si l'on choisit $p_0 = 0{,}05$ alors $\sqrt{\dfrac{p_0(1-p_0)}{n}} = 0{,}0154$ à 10^{-4} près.

Dans ce cas, la zone de rejet de H_0 est la même pour les trois démarches, à savoir [16 ; 200].

En poursuivant la simulation à l'aide d'un tableur, on observe que pour certaines valeurs de p_0, la zone de rejet de H_0 est la même pour les trois démarches et que pour d'autres valeurs de p_0, la zone de rejet de H_0 déterminée par la première démarche contient celle qui est déterminée par les deux autres démarches.

Alors, quelle attitude adopter face à nos étudiants dans le cadre d'un test de conformité d'une proportion ?

Cette présentation peut, par exemple, être faite lors d'une séance de travaux dirigés avec l'outil informatique (usage d'un tableur) ou la calculatrice (avec loi binomiale intégrée), d'autant plus que la démarche mise en œuvre ici est conforme à la méthodologie exposée dans les recommandations pédagogiques qui accompagnent ce programme.