TRAITEMENT GÉNÉRAL D'UN TEST D'HYPOTHÈSE

1 Procédure de décision d'un test d'hypothèse

1. ON ÉCRIT LES DEUX HYPOTHÈSES DU TEST (hypothèse nulle H_0, hypothèse alternative H_1), en français et en langage mathématique.
Pour la formulation mathématique, on introduit ***clairement*** le symbole utilisé. Par ex: m est le salaire moyen des étudiants de première année, p est la proportion d'électeurs qui ont voté pour tel candidat, etc ...
L'hypothèse nulle H_0 est de la forme $p = p_0$, ou $m = m_0$, ... Elle traduit une situation acquise, une hypothèse théorique. L'hypothèse alternative H_1, en revanche, met en question l'hypothèse H_0 et,selon la question posée dans le test, est de la forme
"plus grand que" (ex: $m > m_0$), "plus petit que" (ex: $m < m_0$), "différent de"(ex: $m \neq m_0$).

2. ON FORMULE LA STATISTIQUE DU TEST: sous ***l'hypothèse nulle*** H_0, une certaine *variable aléatoire* X suit une certaine loi de probabilité $\mathcal{L}$. On écrit $X \hookrightarrow \mathcal{L}$.
Par ex: si M_n est la ***moyenne aléatoire*** des salaires (c'est à dire la moyenne des salaires d'un ***échantillon aléatoire*** de taille n), alors M_n suit la loi normale de moyenne 1.2 et d'écart-type 0.4.
On écrit dans cet exemple : $M_n \hookrightarrow \mathcal{N}(1.2, 0.4)$.
Le choix de la variable aléatoire X et de la loi $\mathcal{L}$ dépend du type de test, se fait en s'aidant des dernières pages du formulaire général.

3. ON DÉTERMINE, AU NIVEAU α, LA *RÉGION CRITIQUE* K_α. C'est l'ensemble des valeurs de la variable aléatoire X de l'étape **2.**, qui sont favorables à l'hypothèse alternative de l'étape **1.**, tel que $\mathbb{P}(K_\alpha) = \alpha$. Ici, la probabilité est calculée à l'aide de la *loi de probabilité* déterminée à l'étape **2.**
La forme de la région critique dépend du type de l'hypothèse alternative :
* "plus grand que" : $K_\alpha = [a_\alpha, +\infty[$, "plus petit que" : $K_\alpha =]-\infty, a_\alpha]$,
* "différent de" : $K_\alpha =]-\infty, b_\alpha] \cup [a_\alpha, +\infty[$.

4. DÉCISION DU TEST. L'énoncé décrit une *expérience*, qui permet d'associer une valeur numérique ***expérimentale*** x_e précise à la variable aléatoire X de l'étape **2.** Par exemple, si M_n est la *moyenne aléatoire des âges*, l'échantillon étudié dans l'expérience a une ***moyenne d'âge expérimentale*** m_e de 32 ans.
Si x_0 appartient à la région critique K_α de l'étape **3.** ($x_0 \in K_\alpha$), on ***accepte l'hypothèse*** H_1 au niveau α.
Si x_0 n'appartient pas à K_α, on dit ***qu'on ne peut pas rejeter*** H_0

2 Questions subsidiaires

Signification du test, ou p-value. Si l'expérience donne à la variable aléatoire X une valeur numérique x_0, la p-value est la probabilité (calculée à l'aide de la loi de probabilité de l'étape **2.**) de l'ensemble des valeurs de la variable X favorables à l'hypothèse alternative H_1 et "bordé" par x_0.
* Si H_1 est de type "plus grand que", cet ensemble est $[x_0, \infty[$.
* Si H_1 est de type "plus petit que", cet ensemble est $]-\infty, x_0]$.
Plus x_0 est "franchement" contenu dans la région critique K_α, plus la p-value est petite, et donc plus le test est significatif.

Puissance du test. C'est la probabilité d'accepter l'hypothèse H_1 alors que l'hypothèse H_1 est effectivement satisfaite. On utilise donc la loi de probabilité $\mathcal{L}_1$ de l'étape **2.**, mais dans laquelle le paramètre étudié vérifie l'hypothèse alternative H_1 au lieu de satisfaire l'hypothèse nulle H_0.
Par ex: $p_1 > p_0$ au lieu de p_0, ou $m_1 < m_0$ au lieu de m_0, ... Ce paramètre (p_1, ou m_1, ...) est donné dans l'énoncé.
La puissance est donc la probabilité de la région critique K_α (déterminée à l'étape 3.) avec la nouvelle loi $\mathcal{L}_1$ au lieu de $\mathcal{L}_0$.

VOIR EXEMPLE COMPLET AU VERSO⟹

EXEMPLE D'ÉTUDE D'UN TEST D'HYPOTHÈSE, AJUSTEMENT DE PROPORTION, AVEC LA LOI NORMALE

On estime que seuls 35% des gens ont une aptitude artistique particulièrement développée. Un test psychométrique pratiqué sur 1500 étudiants de différentes écoles de Beaux-Arts pris au hasard, a révélé que 620 étudiants ont obtenus des résultats exceptionnellent bons.

1. Peut-en déduire, au rique d'erreur $\alpha = 5\%$, que les étudiants des Beaux-Arts ont une aptitude artistique particulièrement développée (c'est à dire plus développée que celle de la population générale) ?

2. Quelle est la p-value du test ? 3. Calculer la puissance $\eta\,(40\%)$.

RÉPONSE

1. On pratique un test d'ajustement de proportion. On applique les 4 étapes successives du test.

1. Hypothèses du test.
* Hypothèse nulle H_0: les étudiants des Beaux-Arts n'ont pas une aptitude artistique particulièrement développée.
* Hypothèse alternative H_1 : les étudiants des Beaux-Arts ont une aptitude artistique exceptionnelle.

On note p la proportion d'étudiants d'écoles des Beaux-Arts qui ont un sens artistique exceptionnel, et p_0 la proportion (théorique) de gens qui ont un sens artistique exceptionnel. L'énoncé dit que $p_0 = 35\% = 0.35$.
* Hypothèse nulle : $p = p_0$
* Hypothèse alternative $p > p_0$.

2. On formule la statistique du test. On regarde dans le formulaire le tableau de *test paramétrique* d'ajustement d'une proportion. On a $p_0 = 0.35$ et $q_0 = 0.65$. Ici, $n = 1500$, donc $n > 30$, $n \times p_0 = 525 \geq 5$ et $n \times q_0 = 975 \geq 5$: il s'agit donc d'un grand échantillon. On note P_n la ***proportion aléatoire***. Le tableau dit que, *sous l'hypothèse nulle*, la loi

$$Z = \frac{P_n - p_0}{\sqrt{\frac{p_0 \times q_0}{n}}} \hookrightarrow Z = \mathcal{N}\,(0,1)$$

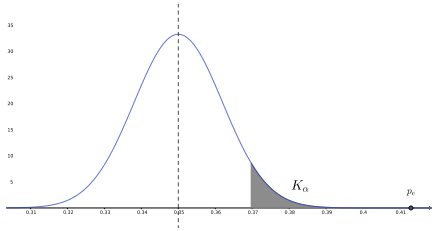
ce qui équivaut à $P_n \hookrightarrow \mathcal{N}\left(p_0, \sqrt{\frac{p_0 \times q_0}{n}}\right)$, c'est à dire $P_n \hookrightarrow \mathcal{N}\,(0.35, 0.012)$

3. Détermination de la région critique K_α, avec $\alpha = 5\% = 0.05$. Puisque l'hypothèse alternative H_1 est de la forme $p > p_0$ (donc du type "plus grand que"), la région critique est de la forme $K_\alpha = P_n \geq p_\alpha$, avec $\mathbb{P}\,[K_\alpha; \mathcal{N}\,(0.35, 0.012)] = 0.05$.
On utilise alors un calcul de ***loi normale inverse***, avec (selon les calculatrices), soit "Tail : Right" et Area:0.05, soit invNorm(0.95,0.35,0.012), et on trouve $p_\alpha = 0.370$.

$$K_\alpha = \{P_n \geq 0.370\}\,.$$

4. Décision du test. La proportion expérimentale est $p_e = \dfrac{620}{1500} = 0.413$. On voit que $p_e > p_\alpha$, et donc que $p_e \in K_\alpha$. Au niveau $\alpha = 5\%$, on admet donc l'hypothèse H_1: les étudiants des Beaux-Arts ont une aptitude artistique exceptionnelle.

2. p-value du test. On calcule la valeur de $\mathbb{P}\,[P_n \geq p_e]$ pour la loi $\mathcal{N}\,(0.35, 0.012)$ de l'étape **2.** On trouve p-value=$\mathbb{P}\,[P_n \geq 0.413; \mathcal{N}\,(0.35, 0.012)] \sim 0$. Le test est donc très significatif !



3. Puissance du test pour $p_1 40\% = 0.40$.
On travaille avec *une autre loi normale :* la loi normale $\mathcal{N}_1$ de moyenne $\mu_1 = 0.40$ et d'écart-type $\sigma_1 = \sqrt{\dfrac{p_1 \times q_1}{n}} = \sqrt{\dfrac{0.40 \times 0.60}{1500}} = 0.013$. Ontrouve

$$\eta\,(40\%) = \mathbb{P}\,[K_\alpha; \mathcal{N}_1] = \mathbb{P}\,[P_n \geq 0.370; \mathcal{N}\,(0.40, 0.013)] = 0.989 \sim 99\%.$$

Si la proportion réelle d'étudiants des Beaux-Arts avec des dons artistiques particuliers est de 40%, le test a 99% de chance d'accepter l'hypothèse alternative H_1 !