EISTI
Statistique 1 : GI

Proposition: Sujet Examen

Exercice ~~1.~~ 2 Dans un procès, on demande au jury de décider entre

l'accusé est innocent : hypothèse H_0

et

l'accusé est coupable : hypothèse H_1.

Pour chacune des 3 questions suivantes, il est proposé plusieurs réponses. Préciser laquelle (ou lesquelles) des affirmations suivantes sont exacte(s) et justifier votre réponse.

1. L'erreur de première espèce (risque α) correspond à:
   a) l'accusé innocent est condamné
   b) l'accusé est relâché faute de preuves
   c) le procès est reporté pour complément d'instruction

2. L'erreur de deuxième espèce (risque β) correspond à:
   a) l'accusé voulant protéger l'auteur du crime, se déclare coupable et il est condamné
   b) un non lieu est prononcé
   c) l'accusé coupable est relâché.

3. On peut envisager deux situations selon la législation du pays:
   - Situation I: tout accusé est supposé innocent jusqu'à preuve de sa culpabilité.
   - Situation II: tout accusé est supposé coupable jusqu'à preuve de son innocence.

   Les risques α et β dans ces deux situations sont tels que :

   a) $\alpha(I) < \alpha(II)$ c) $\beta(I) > \beta(II)$
   b) $\alpha(I) > \alpha(II)$ d) $\beta(I) < \beta(II)$

Réponse

Rappel:

α: probabilité de rejeter à tort H_0
β: probabilité de ne pas rejeter H_0 alors qu'elle est fausse

1.
1 a) est vraie: c'est la définition même du risque α: on rejette H_0 alors qu'elle est vraie

b) aucune décision n'a été prise en ce qui concerne la culpabilité de l'accusé et donc aucun "test" n'a été effectué (donc réponse fausse)

c) on remet à plus tard le "test" et sa conclusion donc cela n'a rien à avoir avec l'un ou l'autre des risques d'erreur (donc réponse fausse).

1 2.

a) Cela antraîne une augmentation du risque α et donc une diminution du risque β (faux)

b) aucune décision n'a été prise quad au caractère coupable ou non; aucun "test" n'a été effectué (faux)

C → d) c'est le non rejet de H_0 alors qu'elle est fausse: c'est la définition de β (vraie) C

(2) 3.

Dans la situation I, il sera plus difficile de prouver la culpabilité de l'accusé que dans la situation II donc on rejettera moins souvent H_0 alors qu'elle fausse $\Rightarrow$ $\beta(I) > \beta(II)$ (d)

Dans la situation II, il sera plus difficile de prouver l'innocence de l'accusé que dans la situation I donc on rejettera plus souvent H_0 alors qu'elle vraie $\Rightarrow$ $\alpha(I) < \cancel{\beta}(II)$ α (a)

Exercice ~~2~~ 3. Pour connaître la proportion p de pharmaciens qui utilisent un système informatique pour effectuer leurs commandes quotidiennes et gérer leurs stocks, un organisme d'étude de marché a effectué un sondage. L'enquête réalisée en 1988 a montré que sur un échantillon de 100 pharmaciens choisies au hasard, 40 utilisent un tel procédé.

1- Donner une estimation ponctuelle $\hat{p}$ et un intervalle de confiance à 95% de p.

2- Quelle doit être la taille de l'échantillon pour connaître p, au risque 5% avec une incertitude absolue inférieure à 1%.

Rappel: Si $C(p) = [p_0 - \Delta(\alpha),\ p_0 + \Delta(\alpha)]$ est un intervalle de confiance à un risque α donné de p , alors la quantité $\Delta(\alpha)$ représente l'imprécision ou l'incertitude absolue de la valeur p.

3- Si l'on prend pour vraie valeur de p la valeur estimé $\hat{p}$ (à la question 1), entre quelles limite se situera la fréquence observée p_0 sur :

un échantiloon de taille	$n = 400$	au risque $\alpha = 5\%$?
un échantiloon de taille	$n = 100$	au risque $\alpha = 1\%$?

Réponse.

1-

$$\hat{p} = \frac{40}{100} = 0.40$$ (0.5)

Nous disposons d'un échantillon assez grand $n = 100$ et la quantité $np(1-p) = 100 \times 0.4 \times 0.6 = 24 > 18$. L'intervalle de confiance à 95% de p est donc égal à :

$$\begin{aligned} IC(95\%) &= \hat{p} \pm z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.40 \pm 1.96\sqrt{\frac{0.4(0.6)}{100}} \\ &= 40\% \pm 10\% \\ &= [\hat{p}_1;\ \hat{p}_2] = [0.304; 0.496]. \end{aligned}$$ (1.5)

2- L'incertitude absolue est de 10%. L'incertitude relative est égale à

$$\frac{\Delta p}{p} = \frac{0.1}{0.4} \times 100 = 25\%$$

Si l'on souhaite une incertitude absolue de 1%, au risque $\alpha = 5\%$, la taille de l'échantillon n sur lequel doit être réalisée l'enquête est tel que :

$$z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.01$$

Le calcul de n, suppose que l'on ait déjà une estimation de p. En prenant comme estimation de p la valeur trouvé dans 1), on obtient :

$$1.96\sqrt{\frac{(0.4)(0.6)}{n}} = 0.01 \Rightarrow n = \frac{(1.96)^2(0.4)(0.6)}{(0.01)^2} = 9220$$ 2

3-

On doit déterminer des intervalles de pari, pour p_0, qu'on note $[P_1; P_2]$ tel que:

$$[P_1; P_2] = p \pm z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

si $n = 400$ et $\alpha = 5\%$ alors $z_\alpha = 1.96$

$$[P_1; P_2] = 0.40 \pm 1.96\sqrt{\frac{(0.4)(0.6)}{400}} = 0.40 \pm 0.05 = [0.35; 0.45]$$ 1,5

si $n = 100$ et $\alpha = 1\%$ alors $z_\alpha = 2.576$

$$[P_1; P_2] = 0.40 \pm 2.576\sqrt{\frac{(0.4)(0.6)}{100}} = 0.40 \pm 0.013 = [0.27; 0.53]$$ 0.5

L'imprécision des estimateurs dépend de la taille de l'échantillon mais aussi du risque d'erreur α.

Exercice ~~3~~ 4. Une compagnie aérienne a demandé une étude statistique afin d'améliorer la sûreté au décollage et de définir un poids limite de bagages. Actuellement, le poids limite des bagages à ne pas dépasser est de 20kg par personne.

Pour l'estimation du poids des voyageurs, sur un échantillon constitué de 400 passagers

$\mathbb{P}(U < 2.99) = 0.9986$ donc $\mathbb{P}(U \geq 2.99) = 0.0014$. On peut donc ...
$\mathbb{P}(U > 4.3) = 0.0014$.

Ainsi, la probababilité qu'un avion de la compagnie ne puisse décoler est quasiment nulle !

6. Sur la base du résultat précédent, il n'est pas nécessaire que la compagnie remette en cause le poids maximum autorisé des bagages de 20kg.

Exercice ~~4~~ 1. On a deux populations I et II. La première a un écart-type égal à ?1 et la deuxième un écart-type égal à ?2=2?1. Pour estimer l?espérance de chacune de ces deux populations, on calcule la moyenne d?un échantillon. On prélève à cette fin un échantillon de chaque population de taille n1 et n2=2n1 respectivement. Pour lequel des deux échantillons l?estimation de la moyenne de la population est la plus précise ? Justifier votre réponse.

Réponses :
Les deux estimateurs (moyenne empirique) sont sans biais. Le plus précis sera donc celui qui a la plus petite variance.

2

$$Var_1 = \frac{\sigma_1^2}{n_1}$$

$$Var_2 = \frac{\sigma_2^2}{n_2} = \frac{4\sigma_1^2}{2n_1} = 2Var_1$$

C'est donc l'estimation correspondant à la première population qui sera la plus précise.

4

Exercice ~~5~~. Une entreprise spécialisée dans la production de confiseries s'engage à respecter certaines normes de fabrication concernant le pourcentage de colorant contenu dans ses produits. Si celui-ci est supérieur ou égal à 3 %, la marchandise est déclarée non conforme et est détruite. A la sortie de la chaîne de fabrication on tire un échantillon aléatoire et indépendant de $N = 9$ mesures en pourcentage. La moyenne et l'écart-type calculés sur cet échantillon sont respectivement $m = 3,3$ % et $s* = 0,2$ %. Le fabricant souhaite savoir s'il peut vendre sa production ou s'il doit la détruire.

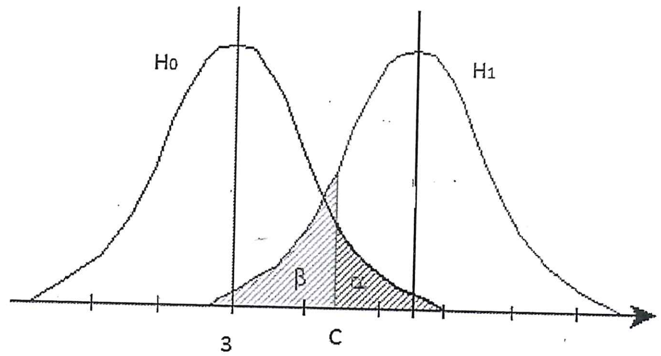
1. Quelle quantité allez-vous tester ? Enoncez les hypothèses du test.
2. Exprimez à l'aide d'une phrase les risques de 1ère et 2ème espèces.
3. Quelle est la statistique (variable de décision) du test ? Quelle est sa loi ? Sous quelle hypothèse ?
4. Représentez graphiquement ce test, en particulier les deux risques et la région critique.
5. Déduire du graphique la forme de la région critique.
6. Calculez le seuil de la région critique pour un risque $\alpha = 1$ %.
7. Pouvez-vous déterminer la puissance du test ?
8. Que conseillez-vous au fabricant, de vendre ou de détruire sa marchandise? Quel risque a-t-il de prendre une mauvaise décision ?

Réponses :

1. On va tester la moyenne de l'échantillon. Les deux hypothèses à départager sont :
   1 (H_0) : $\mu = 3$ %
   (H_1) : $\mu > 3$ %

2. Le risque de première espèce : α est celui de rejeter à tort (H_0).
   Ici c'est le risque de détruire de la marchandise alors que le taux de colorant ne dépasse pas la norme.
   1 Le risque de deuxième espèce, β, est ici le risque de vendre ou de livre de la marchandise avec un taux de colorant excessif.

4.



1

5. On en déduit que la région critique, celle du rejet de (H_0), est de la forme : $\{\overline{X} > C\}$. 1

6. $\alpha = 0,01 = P_{H_0}(\overline{X}_n > C) = P_{H_0}(\dfrac{\overline{X}_n - \mu}{S^*/\sqrt{n}} > \dfrac{C-\mu}{s^*/\sqrt{n}})$.

   La table de la loi de Student, à 8 degrés de liberté, donne alors $\dfrac{C-\mu}{s^*/\sqrt{n}} = 2,8965$. Il faut prendre la colonne correspondant à $0,02$.

   On en déduit : $C = 3 + 2,8965 \times \dfrac{0,2}{3} = 3,19$. 1

7. On ne peut pas calculer la puissance de ce test car l'hypothèse alternative (H_1) est imprécise. Pas de valeur de μ permettant de faire de calcul. 1

8. $\overline{x}_n = 3,3 > C$, on se trouve dans la région critique, et par conséquent on rejette l'hypothèse nulle.
   On doit détruire la marchandise car il y a de fortes chances que le taux de colorant soit trop élevé.
   Le risque pris ici est celui de première espèce : $\alpha = 1$ %. 1