Université de Perpignan - IUT de Carcassonne
Thierry SPINOSA - Vivien ROSSI
Année 2006/2007

IUT STID, $2^{\text{ème}}$ année : **Estimation et tests**

Feuille 2 : Tests Statistiques

Exercice 1

1. Vous vous promenez le long d'une voie ferrée, en pleine rêverie. Vous prenez soudain conscience d'un grand bruit, droit derrière vous. En quoi consiste l'erreur de $1^{\text{ère}}$ espèce (α) ? L'erreur de $2^{\text{ème}}$ espèce. Laquelle est la plus grave ? Que faites-vous ?
2. Vrai ou faux ! ! ! Si on compare le test d'hypothèses à un jugement, on peut dire que l'erreur de $1^{\text{ère}}$ espèce est de condamner un homme coupable.
3. Dans l'analogie du jugement, que se passerait-il, si on souhaitait rendre l'erreur de $1^{\text{ère}}$ espèce nulle ?

Exercice 2

Des relevés effectués pendant de nombreuses années ont permis d'établir que la hauteur annuelle des pluies dans la Beauce (France) [en mm] suit une loi normale $\mathcal{N}(600, 100^2)$.
Des entrepreneurs, surnommés faiseurs de pluie, prétendaient pouvoir augmenter de 50 mm le niveau moyen de pluie, ceci par l'insémination des nuages au moyen d'iodure d'argent et au-delà augmenter le taux de production. Leur procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

Que pouvait-on en conclure ? Deux hypothèses s'affrontaient. Ou bien l'insémination était sans effet, ou bien elle augmentait réellement le niveau moyen de pluie. Si m désigne l'espérance mathématique de la variable aléatoire X égale à la hauteur annuelle de pluie, ces hypothèses peuvent se formaliser comme suit :

$$\begin{cases} \mathrm{H}_0 & : \quad m = 600 \text{ mm (insémination sans effet)} \\ \mathrm{H}_1 & : \quad m > 600 \text{ mm (insémination efficace)} \end{cases}$$

On doit donc tester H_0 (l'hypothèse nulle) contre H_1 (l'hypothèse alternative). Pour cela, on considère la règle de décision suivante :

$$\text{Rejet } \mathrm{H}_0 \iff \bar{X}_9 > 600 + c$$

1. Justifier le choix de cette statistique de test.
2. On considère $c = 50$. Calculer le risque d'erreur de $1^{\text{ère}}$ espèce. Calculer les risques d'erreur de $2^{\text{ème}}$ espèce pour les valeurs 610, 620, 630, 640, 650 et 600 de m. Tracer le graphe de la fonction d'efficacité : $m \mapsto \mathbb{P}(\text{accepter } \mathrm{H}_0 | m)$.
3. Mêmes questions qu'au 2. mais en prenant $c = 55$.
4. Déterminer la constante c telle que le risque d'erreur de $1^{\text{ère}}$ espèce soit égal à 5%. Quelle est la décision à prendre ?

Exercice 3

On considère le problème des faiseurs de pluie de l'exercice 1.

1. On veut tester
$$\begin{cases} \mathrm{H}_0 & : \quad m = 600 \\ \mathrm{H}_1 & : \quad m \neq 600 \end{cases}$$
   (a) Construire le test pour un niveau α donné.

   (b) Déterminer la région critique pour α =5%.

   (c) Calculer les risques d'erreur de 2ème espèce pour les mêmes valeurs de m qu'au 2. de l'exercice 1. Tracer le graphe de la fonction d'efficacité.

2. On veut tester
$$\begin{cases} \mathrm{H}_0 & : \quad m \leq 600 \\ \mathrm{H}_1 & : \quad m > 600 \end{cases}$$
   Montrer que la statistique de test de l'exercice 1 avec pour la constante c la valeur du 4. fournit ici aussi un test de niveau 5%.

Exercice 4

On considère 120 variables aléatoires réelles i.i.d. $X_1, X_2, \ldots, X_{120}$ suivantune loi de Bernoulli de paramètre inconnu p, notée $\mathcal{B}(1, p)$.

1. Construire au niveau α le test
$$\begin{cases} \mathrm{H}_0 & : \quad p = p_0 \\ \mathrm{H}_1 & : \quad p < p_0 \end{cases}$$
2. Préciser le test ci-dessus pour $p_0 = 0.5$ et $\alpha = 5\%$. Calculer alors sa puissance pour $p_1 = 0.25$.
3. On a observé la valeur 24 pour la statistique $\sum_{i=1}^{120} X_i$, donner la degré de significativité de cette réalisation.

Exercice 5

Une machine produit chaque jour un grand nombre de pièces d'un certain type dont la longueur aléatoire (on notera X la variable aléatoire correspondant à la longueur d'une pièce en mm). Pour contrôler la qualité des pièces, on prélève quotidiennement dix pièces au hasard dans la production du jour et on les mesure.

Le responsable de la qualité a adopté les spécifications suivantes :

- si $30.54 \leq X \leq 34.46$, la pièce est conforme
- sinon la pièce est non conforme

De plus, d'après le constructeur de la machine, si la machine est *bien calibrée* la probabilité qu'une pièce produite soit non conforme est de 0.05 (il y a donc en moyenne 5% de pièces non conforme). Si la machine est *mal calibrée* la probabilité qu'une pièce soit non conforme est de 0.15.

1. Quelle est la loi de probabilité de la variable aléatoire M correspondant au nombre de pièces non conformes sur les 10 tirées au hasard dans les deux cas *bien calibrée* et *mal calibrée* ?
2. Calculer la probabilité de l'événement $A = \{M = 0 \ ou \ 1\}$ dans les deux cas *bien calibrée* et *mal calibrée*.
3. On désire extrapoler la qualité des 10 pièces prélevées à la qualité de la production de toute la journée. On adopte la règle de décision suivante :

- si $M = 0$ ou $M = 1$ la production du jour est déclarée acceptable
- sinon la production du jour est déclarée non acceptable et est détruite

(a) Si la machine est bien calibrée, quelle est la probabilité de rejeter la production du jour ?

(b) Si la machine est mal calibrée, quelle est la probabilité de déclarer conforme la production ?

(c) Discuter des résultats obtenus. Comment améliorer la décision, c'est-à-dire diminuer les risques d'erreur ?

Exercice 6

Lors d'une étude sur la fiabilité d'une certaine machine, on s'intéresse au nombre de ses défaillances (noté X) dans une journée. On sait que la loi de probabilité de X est géométrique :

$$\mathbb{P}(X = x) = (1 - m)m^{x-1} \quad x = 1, 2, 3, \ldots$$

où m est un paramètre inconnu $(0 < m < 1)$. Rappelons que pour cette loi $\mathbb{E}(X) = \frac{m}{1-m} = \frac{1}{1-m} - 1$.

A partir de la valeur observée x de X obtenue un jour donné, on souhaite tester les hypothèses suivantes :

$$\begin{cases} H_0 & : \quad m = 0.5 \\ H_1 & : \quad m > 0.5 \end{cases}$$

au moyen du test ci-dessous, où c est une constante arbitraire :

$$\text{Rejet de } H_0 \iff X \geq c$$

1. Interpréter H_1 en l'exprimant en fonction de l'espérance de X.
2. Etant donné le contexte, la forme de la règle de décision vous semble-t-elle adaptée à la forme de H_1 ?
3. Calculer à une précision de 10^{-5} les probabilités $\mathbb{P}(X = x)$ pour $x = 1, \ldots, 10$ avec $m = 0.5$.
4. Calculer c pour que le niveau du test soit voisin de 3%.
5. Calculer la puissance du test obtenu ci-dessus au point $m = 0.9$.
6. Un certain jour on a observé 3 défaillances de la machine. Donner le degré de significativité associé à ce résultat.

Exercice 7

Le nombre de particules émises par une source radioactive par unité de temps en moyenne suit une loi de Poisson. Observant l'emission d'un corps durant 20 unités de temps, on veut déterminer s'il s'agit d'une source radioactive de type A ou de type B. La source de type A emet 0.6 particule par unité de temps et la source de type B 0.8 particules par unité de temps.

1. Préciser les hypothèse que l'on doit tester.
2. Construire le test de niveau α fixé en justifiant le choix de la statistique de test.
3. Déterminer la constante K intervenant dans la région critique pour que le niveau du test soit le plus proche possible de 5%. Calculer sa puissance.

Exercice 8

Dans un atelier une machine fabrique des pièces en grande série. On s'intéresse à leur longueur mesurée en cm. On admet que la variable aléatoire associée à la longueur de chaque pièce suit un loi normale $\mathcal{N}(\mu, \sigma^2 = 0.14^2)$. Afin de tester que la longueur moyenne des pièces $\mu = 150$, on construit un test d'hypothèse. On prélève un échantillon de 49 pièces et on definit la variable aléatoire $\bar{X}_{49}$ comme la moyenne de longueur des 49 pièces.

L'hypothèse nulle est donc $\mu = 150$ et l'hypothèse alternative est $\mu \neq 150$. Le seuil de significativité du test est fixé à 5%.

1. Quelle est sous l'hypothèse nulle H_0, la loi de la variable aléatoire $\bar{X}_n$? Déterminer le réel positif h tel que $\mathbb{P}(150 - h < \bar{X}_n < 150 + h) = 0.95$.
2. Enoncer la règle de décision permettant d'utiliser ce test (i.e. définir le région de rejet)
3. Sur un échantillon de 49 pièces, la moyenne observée est 149.9. Que peut-on conclure au seuil de signification 0.05 ?

Exercice 9

Avant une élection, un candidat affirme qu'il est sûr de dépasser le score des 30%. Pour savoir s'il a raison, on interroge un nombre suffisant de personnes en souhaitant que le risque de dire qu'il a raison alors qu'il se trompe n'excède pas 5%.

1. Quelle est l'hypothèse H_0 ?
2. Quelle est l'hypothèse H_1 ?
3. Quel risque de $1^{\text{ère}}$ espèce choisissez-vous ?
4. Si on interroge 500 personnes, est-ce que 32% d'intentions de votes dans l'échantillon sera considéré comme significatif ?

Exercice 10

Un grand quotidien publie chaque mois la cote de popularité du premier ministre. Le mois précèdent 51% des personnes déclaraient lui faire confiance. Ce moi-ci, à la suite d'un sondage auprès de 1024 personnes, les journaux titres : « Ils ne sont plus que 48% à faire confiance au premier ministre ! » Peut-on affirmer à une erreur de première espèce de 8% puis de 2% que la côte de popularité du premier ministre a baissée ?

Exercice 11

Une entreprise emploie 60 femmes et 40 hommes. Peut-on affirmer, au niveau 4%, que les recruteurs sont sexistes ?

Exercice 12

Un médicament couramment utilisé est connu pour guérir 60% des patients traités. Un nouveau traitement est expérimenté sur 80 patients. On observe 60 guérisons. Ce dernier traitement est-il meilleur, pour un risque de 1% .