PROBABILITES

Contenus :

1. *Estimateur (EMV)*
2. *Intervalle de confiance (IDC)*
3. *Test khi 2*

Bibliographies :

[1] Olivier Gaudoin, Principes et méthodes statistiques

[2] Marc Denier, Probabilités élémentaires, 2006

Intervalle de confiance : (IDC)

1. **IDC pour la moyenne**

Dans [1] , la table 2 de la loi normale donne la valeur u_α telle que $P(|U| > u_\alpha) = \alpha$.

Propriété 5 *Un intervalle de confiance de seuil α pour le paramètre m de la loi $\mathcal{N}(m, \sigma^2)$ est :*

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} u_\alpha, \bar{X}_n + \frac{\sigma}{\sqrt{n}} u_\alpha\right]$$

Le problème est que cet intervalle n'est utilisable que si on connaît la valeur de σ. Or, dans la pratique, on ne connaît jamais les vraies valeurs des paramètres. Une idée naturelle est alors de remplacer σ par un estimateur, par exemple S'_n.

Propriété 6 *Un intervalle de confiance de seuil α pour le paramètre m de la loi $\mathcal{N}(m, \sigma^2)$ est :*

$$\left[\bar{X}_n - \frac{S'_n}{\sqrt{n}} t_{n-1,\alpha}, \bar{X}_n + \frac{S'_n}{\sqrt{n}} t_{n-1,\alpha}\right] = \left[\bar{X}_n - \frac{S_n}{\sqrt{n-1}} t_{n-1,\alpha}, \bar{X}_n + \frac{S_n}{\sqrt{n-1}} t_{n-1,\alpha}\right]$$

Dans l'exemple des niveaux de bruit, on a $n = 20$, $\bar{x}_n = 64.2$ et $s'_n = 5.15$. Pour $\alpha = 5\%$, la table de la loi de Student donne $t_{19,0.05} = 2.093$. On en déduit qu'un intervalle de confiance de seuil 5% pour le niveau de bruit moyen est [61.8, 66.7].

2. **IDC pour la proportion**

Propriété 9 *Un intervalle de confiance asymptotique de seuil α pour p est :*

$$\left[\hat{p}_n - u_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + u_\alpha \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right]$$

Le test du χ^2

Test d'adéquation du χ^2 à une loi entièrement spécifiée

Définition 16 *: On appelle* **test du khi-deux** *le test de H_0 : "les probabilités des k issues sont $p_1, \ldots, p_k$" contre $H_1 = \bar{H}_0$ défini par la région critique :*

$$W = \left\{\sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} > z_{k-1,\alpha}\right\}$$

Dans l'exemple du dé, l'hypothèse que le dé est équilibré s'écrit H_0 : "$\forall j, p_j = \frac{1}{6}$". Alors, la statistique de test vaut $\delta_n^2 = \frac{(42-50)^2}{50} + \ldots + \frac{(61-50)^2}{50} = 6.88$.

Au seuil $\alpha = 5\%$, $z_{5,0.05} = 11.07$. 6.88 est nettement inférieur à 11.07, donc on ne rejette pas H_0 : rien n'indique que le dé n'est pas équilibré.

Exercice 1

Un échantillon de 10 000 personnes sur une population étant donné, on sait que le taux moyen de personnes à soigner pour un problème de cholestérol élevé est de 7,5%. Donner un intervalle dans lequel on soit «sûr» à 95%, de trouver le nombre exact de personnes à soigner sur les 10 000.

Exercice 2

Un vol Marseille - Paris est assuré par un Airbus de 150 places ; pour ce vol des estimations ont montré que la probabilité pour qu'une personne confirme son billet est $p = 0.75$. La compagnie vend n billets, $n > 150$. Soit X la variable aléatoire «nombre de personnes parmi les n possibles, ayant confirmé leur réservation pour ce vol».

1. Quelle est la loi exacte suivie par X ?
2. Quel est le nombre maximum de places que la compagnie peut vendre pour que, à au moins 95%, elle soit sûre que tout le monde puisse monter dans l'avion, c'est-à-dire n tel que : $P[X > 150] \leqslant 0.05$?
3. Reprendre le même exercice avec un avion de capacité de 300 places ; faites varier le paramètre $p = 0.5$; $p = 0.8$.

Exercice 3

Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés ; les observations sur 100 employés tirés au sort sont les suivantes.

taux de cholestérol en cg :(centre classe)	effectif d'employés :
120	9
160	22
200	25
240	21
280	16
320	7

1. Calculer la moyenne m_e et l'écart-type σ_e sur l'échantillon.
2. Estimer la moyenne et l'écart-type pour le taux de cholestérol dans toute l'entreprise.
3. Déterminer un intervalle de confiance pour la moyenne.
4. Déterminer la taille minimum d'échantillon pour que l'amplitude de l'intervalle de confiance soit inférieure à 10.

Exercice 6 – Dans une agence de location de voitures, le patron veut savoir quelles sont les voitures qui n'ont roulé qu'en ville pour les revendre immédiatement. Pour cela, il y a dans chaque voiture une boîte noire qui enregistre le nombre d'heures pendant lesquelles la voiture est restée au point mort, au premier rapport, au deuxième rapport,..., au cinquième rapport. On sait qu'une voiture qui ne roule qu'en ville passe en moyenne 10% de son temps au point mort, 5% en première, 30% en seconde, 30% en troisième, 20% en quatrième, et 5% en cinquième. On décide de faire un test du χ^2 pour savoir si une voiture n'a roulé qu'en ville ou non.

1) Sur une première voiture, on constate sur 2000 heures de conduite : 210 h au point mort, 94 h en première, 564 h en seconde, 630 h en troisième, 390 h en quatrième, et 112 h en cinquième. Cette voiture n'a-t-elle fait que rester en ville ?

2) Avec une autre voiture, on obtient les données suivantes : 220 h au point mort, 80 h en première, 340 h en seconde, 600 h en troisième, 480 h en quatrième et 280 h en cinquième.

Exercice 7 – Une chaîne d'agences immobilières cherche à vérifier que le nombre de biens vendus par agent par mois suit une loi de Poisson de paramètre $\lambda = 1,5$.

1) On observe 52 agents pendant un mois dans la moitié nord de la France. On trouve la répartition suivante : 18 agents n'ont rien vendu, 18 agents ont vendu 1 bien, 8 agents ont vendu 2 biens, 5 agents ont vendu 3 biens, 2 agents ont vendus 4 biens, et un agent a vendu 5 biens. Avec un test du χ^2, chercher s'il s'agit bien de la loi de Poisson attendue.

2) Répondre à la même question avec les 52 agents dans la moitié sud de la France : 19 agents n'ont rien vendu, 20 agents ont vendu un bien, 7 agents 2 biens, 5 agents 3 biens et 1 agent 6 biens.