

**SOCI1241-1 Eléments du calcul des probabilités appliquées aux sciences sociales**  
**Exercices pratiques : chapitre 2**  
**Estimation d'une proportion et comparaison de deux proportions (échantillons indépendants)**

---

**EXERCICES À FAIRE DURANT LA SÉANCE**

- Données : Santé mentale en prison – 2004 – 799 détenus masculins interrogés.
- Variables d'intérêt : antécédents judiciaires (avant 18 ans), personnels et familiaux (pendant l'enfance) des détenus français

**1. Estimation d'une proportion avec a priori non informatif**

- a) Sur base de l'enquête SMP, 333 détenus ont répondu avoir été séparés au moins 6 mois d'un de leurs parents ou de leurs deux parents dans l'enfance et 11 n'ont pas répondu à cette question.
- À partir de là, déterminez si la proportion de détenus masculins en France qui ont connu une séparation parentale dans l'enfance est supérieure à 40%.
  - Déterminez un ensemble de valeurs plausibles pour cette proportion.
- b) Sur base de l'enquête SMP, 220 détenus ont répondu avoir subi des maltraitances dans l'enfance (de nature physique, psychologique ou sexuelle) et 7 n'ont pas répondu à cette question.
- À partir de là, est-il plausible que moins de 30% de la population des détenus ait subi des maltraitances ?
  - Calculez un intervalle de crédibilité pour cette proportion.
- c) Sur base de l'enquête SMP, 220 détenus ont répondu avoir été suivis par un juge pour enfant avant 18 ans et 5 n'ont pas répondu à cette question.
- À partir de là, peut-on dire que la proportion de détenus masculins en France qui ont été suivis par un juge pour enfant avant 18 ans est-elle supérieure à 25%.
  - Déterminez un ensemble de valeurs plausibles pour cette proportion.
- d) Sur base de l'enquête SMP, 181 détenus ont répondu avoir été l'objet d'une mesure de placement (foyer, famille d'accueil) avant 18 ans et 7 n'ont pas répondu à cette question.
- À partir de là, est-il plausible que moins d'un détenu sur quatre ait été l'objet d'une mesure de placement (foyer, famille d'accueil) dans la population d'intérêt ?
  - Calculez un intervalle de crédibilité pour cette proportion.

**2. Estimation d'une proportion avec a priori informatif**

Imaginons à présent (scénario fictif) qu'avant de réaliser l'enquête SMP, une enquête pilote effectuée sur un échantillon aléatoire de 100 détenus ait permis de récolter les informations suivantes :

	Séparation parentale	Maltraitances	Juge pour enfant	Placement
Oui	50	32	37	19
Non	50	68	63	81

Refaites l'exercice 1 en tenant compte des informations issues de l'enquête pilote. Quel impact ces « a priori » ont-ils sur l'estimation des proportions d'intérêt ?

### 3. Comparaison de deux proportions – échantillons indépendants

- a) Le tri croisé des variables « abus » et « séparation » de l'enquête SMP est présenté dans la table ci-dessous :

		Maltraitements	
		Non	Oui
Séparation	Non	376	77
	Oui	190	142

- À partir de ces données, déterminez si la proportion de maltraitements est significativement différente selon qu'il y ait eu ou non une séparation parentale durant l'enfance ?
  - Déterminez un ensemble de valeurs plausibles pour cette différence de proportion.
- b) Peut-on affirmer, sur base des données de l'enquête SMP, qu'un détenu ayant subi des maltraitements durant son enfance a plus de risque de devenir alcoolique plus tard ? Répondez à cette question de deux manières différentes.

		Abus d'alcool	
		Non	Oui
Maltraitements	Non	480	92
	Oui	164	56

- c) Un détenu qui a été placé (foyer, famille d'accueil...) avant ses 18 ans a-t-il significativement plus de « chance » de devenir dépendant (substance) qu'un détenu qui a grandi au sein de son noyau familial ? Répondez à cette question à l'aide d'une probabilité a posteriori adéquate. Fournissez également un intervalle de crédibilité 95% pour quantifier la différence entre les 2 groupes.

		Abus de substances	
		Non	Oui
Placement	Non	473	138
	Oui	109	72

- d) Le fait d'avoir eu des antécédents judiciaires en tant que mineur (via le suivi d'un juge pour enfants) constitue-t-il un facteur de risque de dépression au sein de la population des détenus ? Répondez à cette question à l'aide d'une probabilité a posteriori adéquate. Fournissez également un intervalle de crédibilité 95% pour quantifier la différence entre les 2 groupes.

		Dépression	
		Non	Oui
Juge.enfant	Non	358	216
	Oui	121	99

## EXERCICES COMPLÉMENTAIRES

### Estimation d'une proportion

1. En temps normal dans une usine de production de bois qui possède 250 machines, la probabilité qu'une machine tombe en panne dans une journée de travail ( $\pi$ ) est de 0,01. Un jour, on a observé que parmi 250 appareils, 4 appareils sont tombés en panne.
  - a. Déterminez la densité a posteriori de  $\pi$  connaissant cet événement.
  - b. Déterminez si cet événement indique que la production n'est plus sous contrôle.
  - c. Déterminez un ensemble de valeurs plausibles pour  $\pi$ .
  
2. L'association des fabricants d'automobiles des Etats-Unis relate dans son périodique 'Motor Vehicle Facts and Figures' que, en 1962, parmi toutes les ventes de voitures aux USA, 4.8% sont des voitures étrangères. En 1983, le pourcentage était 27.8%. Du coup, les fabricants américains ont implémenté différentes règles permettant de réduire les coûts de manière à concurrencer les voitures d'importation. Parmi un échantillon de 500 voitures vendues en 1991, 128 sont des voitures d'importation.
  - a. Ces données suggèrent-elles que le pourcentage de voitures étrangères vendues a diminué entre 1983 et 1991 où ce pourcentage valait 27.8%?
  - b. Déterminer un ensemble de valeurs plausibles pour la proportion de voitures étrangères vendues en 1991.
  - c. Supposons que cette proportion de voitures d'importation vendues aux USA soit de nouveau étudiée sur l'année 1992 et que l'on obtienne que parmi un échantillon de 800 voitures, 172 sont des voitures d'importation. Cette nouvelle donnée suggère-t-elle que le pourcentage de voitures étrangères vendues a diminué entre 1983 et 1992 ?
  
3. Il est communément admis que dans une bibliothèque, environ 10% des livres se détériorent chaque année. Le suivi spécifique de 400 livres durant une année a révélé 36 détériorations.
  - a. Que vaut la probabilité (a posteriori) qu'au plus 10% des livres de la bibliothèque ont subi des détériorations cette année-là ?
  - b. Donnez un ensemble de valeurs plausibles pour cette probabilité.

Supposons que l'année suivante sur un échantillon de 300 livres suivis, on sache que 6% de livres aient subi des détériorations.

  - c. En supposant que le pourcentage de détérioration soit stable sur les 2 années, répondez de nouveau aux questions a) et b).
  
4. Supposons que le taux de mortalité, calculé dans un échantillon d'effectif 100 de personnes touchées par une certaine maladie, soit de 0,13. Soit  $\pi$ , cette probabilité de décès au niveau de la population.
  - a. Ces données suggèrent-elles que la probabilité de mourir pour toutes les personnes touchées par cette maladie ( $\pi$ ) soit supérieure à 0,1?
  - b. Répondre à la question en raisonnant à partir d'un ensemble de valeurs plausibles pour cette probabilité.

## Comparaison de deux proportions – échantillons indépendants

5. Un vaccin antigrippal classique est administré à 50 personnes sélectionnées aléatoirement. Dix d'entre elles sont néanmoins grippées l'hiver suivant. Simultanément, un nouveau vaccin est mis sur le marché et testé sur 50 personnes sélectionnées également de façon aléatoire. Cinq parmi celles-ci sont grippées l'hiver suivant. Peut-on affirmer que le nouveau vaccin est plus efficace? Répondez à la question de deux façons différentes.
6. Aux Etats-Unis le « U.S. Energy Information Administration » effectue des études afin d'estimer le pourcentage de ménages américains possédant différents appareils. Les résultats de ces études peuvent être trouvés dans « Residential Energy Consumption Survey: Housing Characteristics ». Euromonitor Publications Limited, London, England conduit des études similaires dans d'autres pays et publie leurs résultats dans « European Marketing Data and Statistics ». Supposons que hors d'un échantillon aléatoire de 500 ménages Américains, 370 possèdent leur propre machine à lessiver et que hors d'un échantillon aléatoire de 450 ménages Français, 365 possèdent cet appareil.
- Ces données suggèrent-elles une différence entre les pourcentages de ménages Américains et Français possédant une machine à lessiver?
  - Donner un ensemble de valeurs plausibles pour cette différence de pourcentages.

### Réponses exercices complémentaires :

#### Exercice 1.

- $(\pi|D) \sim \text{Beta}(5,427) \approx N(0.016; 0.000063)$
- $P(\pi > 0.01|D) = 0.7765 < 0.95 \rightarrow$  non, on ne peut pas dire que la production n'est plus sous contrôle avec les données à disposition.
- IC 95% pour  $\pi$  :  $[0.0004; 0.0315] \ni 0.01$  : même conclusion point b).

#### Exercice 2.

- $P(\pi_{91} < 0.278|D) = 0.8708 < 0.95 \rightarrow$  non, on ne peut pas affirmer que le pourcentage de voitures étrangères vendues a diminué entre 1983 et 1991 avec les données à disposition.
- IC 95% pour  $\pi$  :  $[0.218; 0.294] \ni 0.278$  : même conclusion point a).
- (Indice : a priori informatif pour  $\pi_{92} =$  a posteriori de  $\pi_{91}$ )  
 $P(\pi_{92} < 0.278|D) > 0.9998 > 0.95 \rightarrow$  on conclut que l'affirmation est vraie.

#### Exercice 3.

- $P(\pi < 0.1|D) = 0.7580 < 0.95 \rightarrow$  non, on ne peut pas affirmer que le pourcentage de détérioration soit inférieur à 10% avec les données à disposition.
- IC 95% pour  $\pi$  :  $[0.062; 0.118] \ni 0.1$  : même conclusion point a).
- (Indice : a priori informatif)
- $P(\pi < 0.18|D) = 0.9884 > 0.95 \rightarrow$  cette fois, on peut conclure que le pourcentage de détérioration est inférieur à 10%. IC 95% pour  $\pi$  :  $[0.057; 0.097] < 0.1$  : même conclusion.

#### Exercice 4.

- a)  $P(\pi > 0.1|D) = 0.8133 \rightarrow$  non, on ne peut pas affirmer que la probabilité de mourir dans la population des malades soit supérieure à 0.1 avec les données à disposition.
- b) IC 95% pour  $\pi$  :  $[0.064 ; 0.196] \ni 0.1$  : même conclusion point a).

Exercice 5.

$P(\delta = \pi_2 - \pi_1 < 0|D) = 0.9207 \rightarrow (<0.95)$  non, on ne peut pas affirmer que le nouveau vaccin soit plus efficace avec les données à disposition. IC 95% pour  $\delta = \pi_2 - \pi_1$ :  $[-0.039; 0.0239] \ni 0$  : même conclusion.

Exercice 6.

- a)  $P(\delta = \pi_F - \pi_A > 0|D) = 0.9953 \rightarrow (>0.95)$  oui, la proportion de ménages français possédant une machine à lessiver est significativement supérieure à celle de ménages américains possédant un tel électroménager.
- b) IC 95% pour  $\delta = \pi_F - \pi_A$ :  $[0.017; 0.123] \not\ni 0$  : même conclusion  $\pi_F > \pi_A$