Statistiques Appliquées
Contrôle Continu N°2
TD N°8 et 9

Mathieu VALDENAIRE

Les quatre parties peuvent être traitées de manière indépendante.
Une attention particulière sera portée à la rédaction et à l'explication des résultats.

Partie A : Dépouillement et pronostic sur les résultats d'une élection (5,5 points)

Dans une élection à un tour, 20000 votants sont appelés à se prononcer pour un des deux candidats, John et Hilary.
Les premiers dépouillements indiquent que sur 1% des votants (choisis de manière aléatoire), John recueille 45% des suffrages.

1. Cela signifie-t-il qu'Hilary va l'emporter ? Définir l'outil que vous utilisez pour répondre à cette question et y apporter une réponse, sachant que l'on veut pouvoir accorder 98% de confiance à cette réponse.
2. A partir de quel niveau de confiance pourra-t-on se dire certain de la défaite de John, dans ces conditions ?

Partie B : Nombre d'essais pour remporter une élection (5,5 points)

Le candidat malheureux à l'élection précédente se demande désormais combien de fois il va devoir se présenter avant de remporter une élection.
Ce type de phénomène (nombre d'essais pour faire apparaître un événement de probabilité p), est modélisé par une loi géométrique (on fait l'hypothèse que la probabilité de succès est la même à chaque élection).
La fonction de probabilité de la variable X (nombre d'essais pour remporter une élection), suivant une loi géométrique de paramètre p est la suivante :

$$Pr[X_i = x] = p(1-p)^{x-1}, \quad avec \;\; x = 1, 2, ...$$

Il vous est demandé d'estimer le paramètre de cette loi. On peut pour cela se fonder sur l'observation du temps d'attente de ses prédécesseurs, en supposant que la variable X est distribuée de manière identique. Cette observation, pour ses douze prédécesseurs, aboutit à l'échantillon :

1 1 6 4 1 2 3 2 1 4 2 1

1. Donner une estimation du paramètre de cette loi par la méthode des moments. L'espérance d'une loi géométrique est $E(X) = 1/p$ et sa variance $V(X) = (1-p)/p^2$).
2. Calculer l'estimateur du maximum de vraisemblance de ce paramètre et en déduire une estimation du paramètre dans le cas présent. Commenter.

Partie C : Prévision du nombre de voix à la prochaine élection (4 points)

Le candidat malheureux à la première élection souhaite désormais prévoir le nombre de voix qu'il obtiendra à la prochaine élection (variable Y).

On retient désormais l'approximation de la loi de la variable Y par la loi normale pour la suite de l'exercice. On estime les paramètres de la loi de cette variable Y à partir de l'observation du score des candidats du même parti aux précédentes élections (70 observations), en supposant qu'à chaque élection le nombre de leurs voix est indépendant et identiquement distribué. La proportion d'électeurs votant pour le candidat en question est ainsi estimée à 48% des voix, soit 9600 voix, et la variance estimée est égale à 4992.
Entre quelles bornes le pourcentage de voix de ce candidat à la prochaine élection se situera-t-il ? (à 95% de confiance) ?

Note. L'espérance d'une variable X suivant une loi de Bernouilli est $E(X) = p$ et sa variance $V(X) = p(1-p)$.

Partie D : Modélisation économétrique du nombre de voix aux élections (5 points)

Les modèles les plus performants de prévision électorale font dépendre le nombre de voix de la majorité sortante au second tour d'une élection des variables suivantes :

$$VOT_{it} = c_i + \alpha_1 DPIB_t + \alpha_2 CHO_{it} + \alpha_3 POP_t + \alpha_4 PRE_{it} + \alpha_5 VP_{it} + \alpha_6 ELI_{it} + \epsilon_{it}$$

où la variable dépendante VOT représente le nombre de voix obtenu à l'élection, et les variables explicatives retenues sont $DPIB_t$, la différence entre le taux de croissance réelle du PIB l'année des élections et celui de l'année précédant l'année des élections, POP_t la popularité du Premier ministre, PRE_{it} est une variable politique constituée par les résultats des élections précédentes, VP_{it} une variable reflétant les différences partisanes entre le département où a lieu le vote et la moyenne au niveau national, et enfin une variable ELI_{it} qui reflète la probabilité pour la majorité sortante d'avoir été éliminée dès le premier tour.

1. Quel est l'intérêt d'un tel modèle pour prévoir le résultat des élections ? (1 point)
2. Quelle est la signification du coefficient α_1 ? Quel signe peut-on attendre pour ce coefficient ? (1 point)

   L'estimation de ce modèle par la méthode des moindres carrés ordinaires (MCO) sur les données des élections législatives sur la période 1981-1997 donne les résultats suivants :

   $$VOT_{it} = c_i + 0.91DPIB_t + 4.76POP_t + 0.54PRE_{it} + 0.30VP_{it} - 0.29ELI_{it} + e_{it}$$

   (0.210) (0.889) (0.040) (0.037) (0.012)
   (les chiffres entre parenthèses sont les écart-types estimés des coefficients correspondants de l'équation ci-dessus).

3. Pourquoi le coefficient d'une régression de MCO est-il une variable aléatoire ? (0,5 point)
4. Rappeler les propriétés des estimateurs des MCO (1 point)
5. La popularité du premier ministre, reflétée par la variable POP_t a-t-elle un impact significatif sur le nombre de voix obtenu par la majorité sortante aux élections ?(1.5 point)