

Feuille de Travaux Dirigés n° 3

Choix de modèle

Ces exercices sont issus du livre : « Analyse de régression appliquée » de Yadolah Dodge, Édition Dunod.

Exercice III.1. Le tableau au verso présente pour 48 états des États-Unis les quantités suivantes :

- TAX : taxe sur le carburant en cents par gallon en 1972;
- DLIC : pourcentage de la population qui possède un permis de conduire;
- INC : revenu par tête en milliers de dollars en 1972;
- ROAD : milliers de miles d'autoroutes recevant l'aide fédérale en 1971;
- FUEL : consommation de carburant par tête.

1. Trouver le meilleur modèle pour expliquer la consommation de carburant à l'aide de la méthode stepwise.
2. Trouver le meilleur modèle à l'aide de la méthode stagewise.

Consommation de carburant dans les états américains.

Source : Weisberg (1985).

| Observation | État | TAX | DLIC | INC | ROAD | FUEL |
|-------------|------|----------|----------|----------|----------|----------|
| i | | x_{i1} | x_{i2} | x_{i3} | x_{i4} | x_{i5} |
| 1 | ME | 9 | 52,5 | 3,571 | 1,976 | 541 |
| 2 | NH | 9 | 57,2 | 4,092 | 1,250 | 524 |
| 3 | VT | 9 | 58,0 | 3,865 | 1,586 | 561 |
| 4 | MA | 7,50 | 52,9 | 4,870 | 2,351 | 414 |
| 5 | RI | 8 | 54,4 | 4,399 | 0,431 | 410 |
| 6 | CN | 10 | 57,1 | 5,342 | 1,333 | 457 |
| 7 | NY | 8 | 45,1 | 5,319 | 11,868 | 344 |
| 8 | NJ | 8 | 55,3 | 5,126 | 2,138 | 467 |
| 9 | PA | 8 | 52,9 | 4,447 | 8,577 | 464 |
| 10 | OH | 7 | 55,2 | 4,512 | 8,507 | 498 |
| 11 | IN | 8 | 53,0 | 4,391 | 5,939 | 580 |
| 12 | IL | 7,50 | 52,5 | 5,126 | 14,186 | 471 |
| 13 | MI | 7 | 57,4 | 4,817 | 6,900 | 525 |
| 14 | WI | 7 | 54,5 | 4,207 | 6,580 | 508 |
| 15 | MN | 7 | 60,8 | 4,332 | 8,159 | 566 |
| 16 | IA | 7 | 58,6 | 4,318 | 10,340 | 635 |
| 17 | MO | 7 | 57,2 | 4,206 | 8,508 | 603 |
| 18 | ND | 7 | 54,0 | 3,718 | 4,725 | 714 |
| 19 | SD | 7 | 72,4 | 4,716 | 5,915 | 865 |
| 20 | NE | 8,50 | 67,7 | 4,341 | 6,010 | 640 |
| 21 | KS | 7 | 66,3 | 4,593 | 7,834 | 649 |
| 22 | DE | 8 | 60,2 | 4,983 | 0,602 | 540 |
| 23 | MD | 9 | 51,1 | 4,897 | 2,449 | 464 |
| 24 | VA | 9 | 51,7 | 4,258 | 4,686 | 547 |
| 25 | WV | 8,50 | 55,1 | 4,574 | 2,619 | 460 |
| 26 | NC | 9 | 54,4 | 3,721 | 4,746 | 566 |
| 27 | SC | 8 | 54,8 | 3,448 | 5,399 | 577 |
| 28 | GA | 7,50 | 57,9 | 3,846 | 9,061 | 631 |
| 29 | FL | 8 | 56,3 | 4,188 | 5,975 | 574 |
| 30 | KY | 9 | 49,3 | 3,601 | 4,650 | 534 |
| 31 | TN | 7 | 51,8 | 3,640 | 6,905 | 571 |
| 32 | AL | 7 | 51,3 | 3,333 | 6,594 | 554 |
| 33 | MS | 8 | 57,8 | 3,063 | 6,524 | 577 |
| 34 | AR | 7,50 | 54,7 | 3,357 | 4,121 | 628 |
| 35 | LA | 8 | 48,7 | 3,528 | 3,495 | 487 |
| 36 | OK | 6,58 | 62,9 | 3,802 | 7,834 | 644 |
| 37 | TX | 5 | 56,6 | 4,045 | 17,782 | 640 |
| 38 | MT | 7 | 58,6 | 3,897 | 6,385 | 704 |
| 39 | ID | 8,50 | 66,3 | 3,635 | 3,274 | 648 |
| 40 | WY | 7 | 67,2 | 4,345 | 3,905 | 968 |
| 41 | CO | 7 | 62,6 | 4,449 | 4,639 | 587 |
| 42 | NM | 7 | 56,3 | 3,656 | 3,985 | 699 |
| 43 | AZ | 7 | 60,3 | 4,300 | 3,635 | 632 |
| 44 | UT | 7 | 50,8 | 3,745 | 2,611 | 591 |
| 45 | NV | 6 | 57,2 | 5,215 | 2,302 | 782 |
| 46 | WN | 9 | 57,1 | 4,476 | 3,942 | 510 |
| 47 | OR | 7 | 62,3 | 4,296 | 4,083 | 610 |
| 48 | CA | 7 | 59,3 | 5,002 | 9,794 | 524 |

Exercice III.2. L'ensemble des données du tableau au verso a été créé suite à une étude réalisée en 1976 sur la qualité de l'eau pour les rivières de l'état de New York. La concentration en azote a été utilisée comme indicateur de la qualité de l'eau dans les 20 rivières suivantes :

- Olean
- Oatka
- Hackensack
- Fishkill
- Susquehanna
- Tioughnioga
- East Canada
- Ausable
- Schoharie
- Oswegatchie
- Cassadaga
- Neversink
- Wappinger
- Honeoye
- Chenango
- West Canada
- Saranac
- Black
- Raquette
- Cochocton.

Les variables utilisées sont :

- X_1 : Agriculture : pourcentage de terres cultivées
- X_2 : Forêt : pourcentage de forêts
- X_3 : Résidence : pourcentage de terres en zone résidentielle
- X_4 : Commerce et industrie : pourcentage en terres en zone commerciale ou industrielle
- Y : Quantité d'azote : concentration moyenne (mg/L) basée sur des échantillons prélevés à intervalles réguliers durant le printemps et l'été.

1. Trouver le meilleur modèle pour expliquer la quantité d'azote dans les rivières de l'état de New York.
2. Peut-on améliorer le résultat trouvé à la question 1. en enlevant certaines observations ? Justifier.

Qualité de l'eau des rivières de l'état de New York

Source : A. Haith (1976)

| Observation | Agriculture | Forêts | Habitations | Com. et ind. | Azote |
|-------------|-------------|----------|-------------|--------------|-------|
| i | x_{i1} | x_{i2} | x_{i3} | x_{i4} | y |
| 1 | 26 | 63 | 1, 2 | 0, 29 | 1, 10 |
| 2 | 29 | 57 | 0, 7 | 0, 09 | 1, 01 |
| 3 | 54 | 26 | 1, 8 | 0, 58 | 1, 90 |
| 4 | 2 | 84 | 1, 9 | 1, 98 | 1, 00 |
| 5 | 3 | 27 | 29, 4 | 3, 11 | 1, 99 |
| 6 | 19 | 61 | 3, 4 | 0, 56 | 1, 42 |
| 7 | 16 | 60 | 5, 6 | 1, 11 | 2, 04 |
| 8 | 40 | 43 | 1, 3 | 0, 24 | 1, 65 |
| 9 | 28 | 62 | 1, 1 | 0, 15 | 1, 01 |
| 10 | 26 | 60 | 0, 9 | 0, 23 | 1, 21 |
| 11 | 26 | 53 | 0, 9 | 0, 18 | 1, 33 |
| 12 | 15 | 75 | 0, 7 | 0, 16 | 0, 75 |
| 13 | 6 | 84 | 0, 5 | 0, 12 | 0, 73 |
| 14 | 3 | 81 | 0, 8 | 0, 35 | 0, 80 |
| 15 | 2 | 89 | 0, 7 | 0, 35 | 0, 76 |
| 16 | 6 | 82 | 0, 5 | 0, 15 | 0, 87 |
| 17 | 22 | 70 | 0, 9 | 0, 22 | 0, 80 |
| 18 | 4 | 75 | 0, 4 | 0, 18 | 0, 87 |
| 19 | 21 | 56 | 0, 5 | 0, 13 | 0, 66 |
| 20 | 40 | 49 | 1, 1 | 0, 13 | 1, 25 |

Exercice III.3. Le tableau au verso présente un ensemble de données avec 13 variables pour expliquer le taux d'accidents (Y) dans l'état du Minnesota. Les données comprennent 39 observations faites sur des tronçons d'autoroute. Les

variables retenues sont les suivantes :

- X_1 : longueur du tronçon (en miles) ;
- X_2 : trafic moyen quotidien (en milliers de véhicules) ;
- X_3 : pourcentage du volume de camions par rapport au volume total ;
- X_4 : vitesse limitée autorisée (en miles par heure) ;
- X_5 : largeur de la piste (en pieds) ;
- X_6 : largeur de la piste d'arrêt d'urgence (en pieds) ;
- X_7 : nombre de changements de pistes libres (par mile sur le tronçon) ;
- X_8 : nombre de changements de pistes signalés (par mile) ;
- X_9 : nombre de points d'entrée sur l'autoroute (par mile sur le tronçon) ;
- X_{10} : nombre total de pistes (dans les deux directions) ;
- X_{11} : 1 s'il s'agit d'une autoroute fédérale inter-état, 0 sinon ;
- X_{12} : 1 s'il s'agit d'une artère principale d'autoroute, 0 sinon ;
- X_{13} : 1 s'il s'agit d'une artère majeure d'autoroute, 0 sinon.

À l'aide d'une des méthodes de sélection de variables présentées dans le cours, choisir le meilleur modèle pour expliquer le taux d'accidents (Y) dans l'état du Minnesota. Justifier votre choix.

| i | $x_{i,1}$ | $x_{i,2}$ | $x_{i,3}$ | $x_{i,4}$ | $x_{i,5}$ | $x_{i,6}$ | $x_{i,7}$ | $x_{i,8}$ | $x_{i,9}$ | $x_{i,10}$ | $x_{i,11}$ | $x_{i,12}$ | $x_{i,13}$ | y_i |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|------------|-------|
| 1 | 4,99 | 69 | 8 | 55 | 12 | 10 | 1,20 | 0,00 | 4,6 | 8 | 1 | 0 | 0 | 4,58 |
| 2 | 16,11 | 73 | 8 | 60 | 12 | 10 | 1,43 | 0,00 | 4,4 | 4 | 1 | 0 | 0 | 2,86 |
| 3 | 9,75 | 49 | 10 | 60 | 12 | 10 | 1,54 | 0,00 | 4,7 | 4 | 1 | 0 | 0 | 3,02 |
| 4 | 1,65 | 61 | 13 | 65 | 12 | 10 | 0,94 | 0,00 | 3,8 | 6 | 1 | 0 | 0 | 2,29 |
| 5 | 20,01 | 28 | 12 | 70 | 12 | 10 | 0,65 | 0,00 | 2,2 | 4 | 1 | 0 | 0 | 1,61 |
| 6 | 5,97 | 30 | 6 | 55 | 12 | 10 | 0,34 | 1,84 | 24,8 | 4 | 0 | 1 | 0 | 6,87 |
| 7 | 8,57 | 46 | 8 | 55 | 12 | 8 | 0,47 | 0,70 | 11,0 | 4 | 0 | 1 | 0 | 3,85 |
| 8 | 5,24 | 25 | 9 | 55 | 12 | 10 | 0,38 | 0,38 | 18,5 | 4 | 0 | 1 | 0 | 6,12 |
| 9 | 15,79 | 43 | 12 | 50 | 12 | 4 | 0,95 | 1,39 | 7,5 | 4 | 0 | 1 | 0 | 3,29 |
| 10 | 8,26 | 23 | 7 | 50 | 12 | 5 | 0,12 | 1,21 | 8,2 | 4 | 0 | 1 | 0 | 5,88 |
| 11 | 7,03 | 23 | 6 | 60 | 12 | 10 | 0,29 | 1,85 | 5,4 | 4 | 0 | 1 | 0 | 4,20 |
| 12 | 13,28 | 20 | 9 | 50 | 12 | 2 | 0,15 | 1,21 | 11,2 | 4 | 0 | 1 | 0 | 4,61 |
| 13 | 5,40 | 18 | 14 | 50 | 12 | 8 | 0,00 | 0,56 | 15,2 | 2 | 0 | 1 | 0 | 4,80 |
| 14 | 2,96 | 21 | 8 | 60 | 12 | 10 | 0,34 | 0,00 | 5,4 | 4 | 0 | 1 | 0 | 3,85 |
| 15 | 11,75 | 27 | 7 | 55 | 12 | 10 | 0,26 | 0,60 | 7,9 | 4 | 0 | 1 | 0 | 2,69 |
| 16 | 8,86 | 22 | 9 | 60 | 12 | 10 | 0,68 | 0,00 | 3,2 | 4 | 0 | 1 | 0 | 1,99 |
| 17 | 9,78 | 19 | 9 | 60 | 12 | 10 | 0,20 | 0,10 | 11,0 | 4 | 0 | 1 | 0 | 2,01 |
| 18 | 5,49 | 9 | 11 | 50 | 12 | 6 | 0,18 | 0,18 | 8,9 | 2 | 0 | 1 | 0 | 4,22 |
| 19 | 8,63 | 12 | 8 | 55 | 13 | 6 | 0,14 | 0,00 | 12,4 | 2 | 0 | 1 | 0 | 2,76 |
| 20 | 20,31 | 12 | 7 | 60 | 12 | 10 | 0,05 | 0,99 | 7,8 | 4 | 0 | 1 | 0 | 2,55 |
| 21 | 40,09 | 15 | 13 | 55 | 12 | 8 | 0,00 | 0,12 | 9,6 | 4 | 0 | 1 | 0 | 1,89 |
| 22 | 11,81 | 8 | 8 | 60 | 12 | 10 | 0,00 | 0,00 | 4,3 | 2 | 0 | 1 | 0 | 2,34 |
| 23 | 11,39 | 5 | 9 | 50 | 12 | 8 | 0,00 | 0,09 | 11,1 | 2 | 0 | 1 | 0 | 2,83 |
| 24 | 22,00 | 5 | 15 | 60 | 12 | 7 | 0,56 | 0,00 | 6,8 | 2 | 0 | 1 | 0 | 1,81 |
| 25 | 3,58 | 23 | 6 | 40 | 12 | 2 | 0,31 | 2,51 | 53,0 | 4 | 0 | 0 | 1 | 9,23 |
| 26 | 3,23 | 13 | 6 | 45 | 12 | 2 | 0,13 | 0,93 | 17,3 | 2 | 0 | 0 | 1 | 8,60 |
| 27 | 7,73 | 7 | 8 | 55 | 12 | 8 | 0,00 | 0,52 | 27,3 | 2 | 0 | 0 | 1 | 8,21 |
| 28 | 14,41 | 10 | 10 | 55 | 12 | 6 | 0,09 | 0,07 | 18,0 | 2 | 0 | 0 | 1 | 2,93 |
| 29 | 11,54 | 12 | 7 | 45 | 12 | 3 | 0,00 | 0,09 | 30,2 | 2 | 0 | 0 | 1 | 7,48 |
| 30 | 11,10 | 9 | 8 | 60 | 12 | 7 | 0,00 | 0,00 | 10,3 | 2 | 0 | 0 | 1 | 2,57 |
| 31 | 22,09 | 4 | 8 | 45 | 11 | 3 | 0,00 | 0,14 | 18,2 | 2 | 0 | 0 | 1 | 5,77 |
| 32 | 9,39 | 5 | 10 | 55 | 13 | 1 | 0,00 | 0,00 | 12,3 | 2 | 0 | 0 | 1 | 2,90 |