

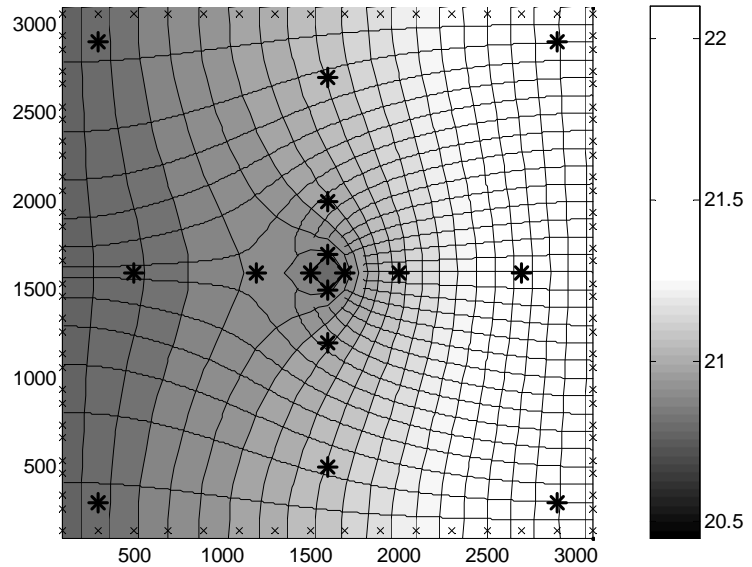
Exemple d'application du test d'ajout

Données : charges hydrauliques observées à 16 piézomètres.

Variables : coordonnées à x, y et la charge h(x,y). les données ont été obtenus à l'aide du simulateur d'écoulement SEEP-W. (charges imposées aux frontières E-O; frontières N-S imperméables).

Autre information : un puits à fort débit se trouve en (x,y)=(1600,1600). On calcule la distance² entre chaque observation et le puits (d²).

x	y	h
1600	1700	21.283
1700	1600	21.316
1600	1500	21.283
1500	1600	21.249
1600	2000	21.44
2000	1600	21.573
1600	1200	21.44
1200	1600	21.306
1600	2700	21.554
2700	1600	21.921
1600	500	21.554
500	1600	21.188
2900	2900	22.046
2900	300	22.046
300	300	21.179
300	2900	21.179



$$\bar{h} = 21.472$$

$$\sum_i h_i^2 = 7378.2$$

On considère les modèles suivants :

Modèle	Équation	SCE	R ²
A	$h=b_0+b_1x+e$	0.253	0.807
B	$h= b_0+b_1x+b_2y+e$	0.253	0.807
C	$h= b_0+b_1x+b_2d^2+e$	0.077	0.9415
D	$h= b_0+b_1x +b_2y+b_3x^2+b_4xy+b_5y^2+e$	0.077	0.9415
E	$h= b_0+b_1x+b_2y+b_3\ln(d^2)+e$	~0	~1
F	$h= b_0+b_1x+ b_2(\text{revenus annuels de 16 étudiants de la classe})+b_3(\text{poids des 16 étudiants})+b_4(\text{taille des 16})\dots+b_{15}(\text{heures de sommeil au cours de la dernière semaine})+e$	0	1
G	Modèle F sans la 15 ^e variable	0.0011	0.999

Questions :

- 1- Que vaut SCT ? SCTm ? SCE si on utilise un modèle avec seulement la constante ?
- 2- Le modèle A est-il significatif ?
- 3- Le modèle B ajoute-t-il de l'information ? Que concluez-vous ?
- 4- Le modèle C ajoute-t-il de l'information (par rapport à A) ?
- 5- Avez-vous une explication pour le fait que les statistiques soient identiques pour les modèles C et D ?
- 6- L'équation de la charge en fonction de la distance à un puits, en régime permanent et pour un aquifère infini est :

$$h(x, y) = h_w - i_x \Delta x - i_y \Delta y - \frac{Q}{4\pi T} \ln \left(\frac{r^2}{r_w^2} \right)$$

Comparez avec le modèle E. Exprimez les gradients en x et y et la transmissivité T en fonction des coefficients de la régression.

- 7- Que pensez-vous du modèle F ? Peut-on tester l'ajout des variables 2 à 15 ?
- 8- L'ajout du modèle G par rapport au modèle A est-il significatif ?

Corrigé :

1- $SCT=7378.2$

$SCTm=7378.2-16*21.472^2=1.312$

2- test ajout : $\frac{(1.312 - 0.253)/1}{0.253/14} = 58.6$ Fortement significatif $F(1,14;0.05)=4.6$

3- test=0; Non-significatif. Le gradient régional est // à x.

4- test : $\frac{(0.253 - 0.077)/1}{0.077/13} = 29.7$ Fortement significatif $F(1,13;0.05)=4.67$

5- La composante quadratique en d^2 peut être prise en charge complètement par les termes linéaires et quadratiques du modèle D. Le puits est réellement l'élément contrôlant la forme de la charge, sinon on aurait eu un R^2 plus grand avec D qu'avec C.

6- $ix=-b_1$; $iy=-b_2$; $T=-Q/(4\pi b_3)$. Les termes r_w et h_w sont amalgamés dans la constante b_0 .

7- Pas fort comme modèle même si le R^2 vaut 1 ! N'importe quel ensemble de 15 variables permettra d'obtenir ce résultat. Le hic, c'est que si l'on mesurait une 17^e charge, le modèle serait probablement très mauvais pour prédire celle-ci. On ne peut tester l'ajout car l'on a alors une division par 0.

8- Le R^2 est très fort mais le modèle est idiot. Le test d'ajout donne $\frac{(0.253 - 0.0011)/13}{0.0011/1} = 16.14$. Cette valeur est très faible par rapport à une $F(13,1;0.05)=244.7$. Clairement on doit conclure que l'ajout est non-significatif.

Note: une autre façon de vérifier la validité du modèle est d'exclure une donnée à tour de rôle, effectuer la régression sans cette donnée et calculer la valeur prédite pour l'observation qui a été exclue avec le modèle obtenu. On obtient ainsi une erreur pour chaque observation qui n'est pas influencée par l'observation elle-même. Avec cette stratégie, on calcule par exemple pour le modèle E un « R^2 » de 1. Pour le modèle G, le « R^2 » (calculé avec $1-SCE/SCTm$) vaut alors -2.3 au lieu de 0.999. Bref, le modèle E a une excellente capacité de prédiction alors que le modèle G est en fait beaucoup moins efficace que la simple moyenne de Y. Pour sa part le modèle F montre alors un R^2 de -223 .