QUESTION 1 (20 points)

Afin de comparer la résistance à la traction (en *lb/po²*) de deux types de câbles, on prélève deux échantillons (un par type) de taille 16 chacun. Les calculs effectués sur les données des deux échantillons donnent les résultats qu'on résume dans le tableau suivant :

	Taille de l'échantillon	Moyenne	Variance
Type 1	16	45,52	11,2
Type 2	16	48,86	14,3

On suppose que la résistance à la traction des câbles de type 1 et type 2 suivent respectivement les lois normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$.

a) Construire un intervalle de confiance bilatéral, au niveau 95%, pour le rapport $\frac{\sigma_1^2}{\sigma_2^2}$. Que peut-on conclure ? ***[0.273, 2.24]***

b) Y a-t-il une raison de croire que le type 2 a, en moyenne, une résistance à la traction supérieure à celle du type 1 ? Répondre à cette question à l'aide d'un test d'hypothèses au seuil 5%. ***Oui car H_0 rejetée donc $\boldsymbol{\mu_2}$ plus grand que $\boldsymbol{\mu_1}$***

c) Calculer la puissance du test effectué en b) si la résistance moyenne réelle du type 2 est supérieure de 3 *lb/po²* à celle du type 1. On suppose dans cette question <u>seulement</u> que les vraies variances des résistances à la traction des câbles de type 1 et 2 sont $\sigma_1^2 = \sigma_2^2 = 12{,}75$.
$\mathbf{1 - \beta = 0.76}$

d) Le fabricant des câbles de type 2 prétend que la différence entre les deux moyennes théoriques $(\mu_2 - \mu_1)$ est supérieure à 3 *lb/po²*. Effectuer le test d'hypothèses approprié (au seuil 5%) afin de déterminer si ce fabricant a raison. ***Accepter H_0, le fabricant n'a pas raison***

<u>Aide</u> : on donne $F_{14,14}(0{,}025) = 2{,}98$, $F_{15,15}(0{,}025) = 2{,}86$, $F_{16,16}(0{,}025) = 2{,}76$.

<u>Réponse :</u>

QUESTION 2 (12 points)

Un ingénieur géologue fournit les informations suivantes sur un gisement de cuivre. La variable étudiée est la teneur en cuivre (exprimée en %) :

- Les teneurs varient 10 fois plus vite verticalement qu'horizontalement;
- dans le plan horizontal, on ne dénote pas de directions géologiques préférentielles;
- dans le plan horizontal on observe une corrélation nulle entre les teneurs de points espacés de 200 m ou plus. Pour des distances inférieures à 200 m, on note toutefois une corrélation qui va croissante au fur et à mesure que les observations se rapprochent;
- en fonction de la distance, la corrélation entre les teneurs varie assez linéairement, du moins sur les 130 premiers mètres;
- une série d'analyse de quelques paires d'observations spatialement rapprochées a été effectuée. La variance des différences de teneurs des paires d'observations a donné la valeur 2 $\%^2$;
- la variance des teneurs des observations prises sur l'ensemble du gisement est de 7 $\%^2$.

a) Décrire un modèle de variogramme 3D permettant de tenir compte simultanément de toutes ces informations fournies par l'ingénieur (donner le type du modèle et ses paramètres).

Portée horizontale 200m
Portée verticale 20m
Effet de pépite 1$\%^2$
Palier=variance=7$\%^2$

b) Calculer la covariance entre les teneurs en cuivre observées en deux points situés dans le plan horizontal et séparés par une distance de 100 m. ***1.875$\%^2$***

Réponse :

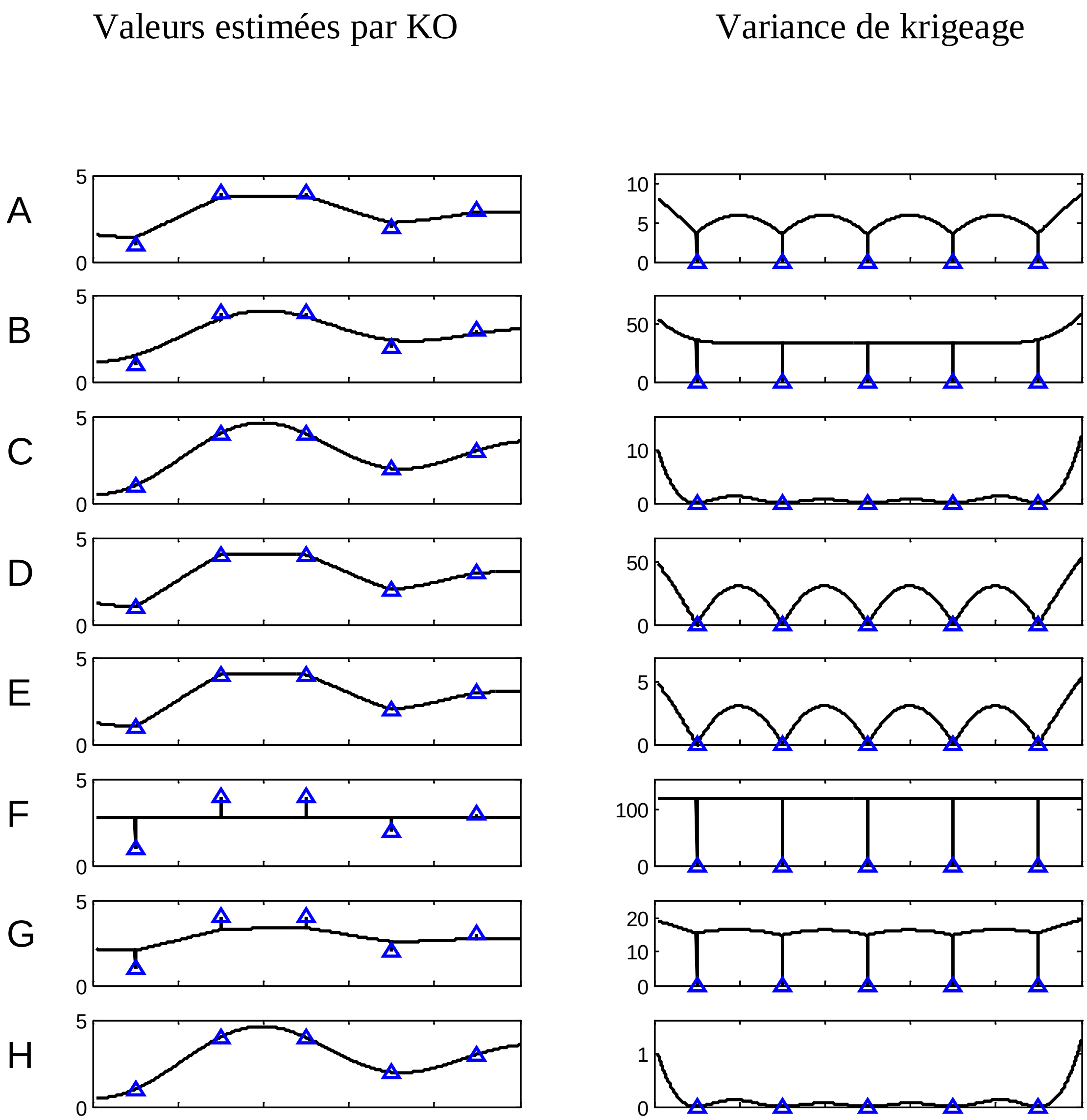
QUESTION 3 (10 points)

Le tableau suivant présente 8 modèles de variogramme. La figure de la page suivante montre les résultats d'un krigeage en 1D effectué avec les mêmes points échantillons et chacun des modèles du tableau. Les figures de gauche donnent la valeur estimée en fonction de la coordonnée x et les figures de droite donnent la variance de krigeage en fonction de la coordonnée x.

Associez une figure (de A à H) à chacun des modèles du tableau (répondez dans le tableau). Chaque figure comporte d'abord les valeurs estimées à gauche et les variances de krigeage correspondantes à droite.

Modèle #	Description	Figure associée (A à H)
1	Sphérique a = 50, C =10	
2	Sphérique a = 50, C = 100	
3	Sphérique a=50, C_0=10, C=10	
4	Sphérique, a=50, C_0 = 2, C = 10	
5	Gaussien $a_{effectif}$ = 50, C = 10	
6	Gaussien $a_{effectif}$ = 50, C_0=20, C = 100	
7	Gaussien $a_{effectif}$ = 50, C = 100	
8	Effet de pépite pur avec C_0 = 100	

***Rép :** dans l'ordre E-D-G-A-H-B-C-F*


Valeurs estimées par KO
Variance de krigeage
A
B
C
D
E
F
G
H
5
0
10
50
100
20
1

QUESTION 4 (22 points)

On envisage une analyse de régression pour établir la meilleure équation de prédiction d'une variable Y en utilisant trois variables explicatives X_1, X_2 et X_3. On dispose pour cela de 24 observations de la forme $(X_{i1}, X_{i2}, X_{i3}, Y_i)$, $i = 1, 2, \ldots, 24$.

Le modèle considéré est :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

où β_0, β_1, β_2 et β_3 sont des paramètres et ε est une erreur aléatoire.

Suite à l'ajustement du modèle, on dispose des résultats partiels suivants :

$$\sum_{i=1}^{24} Y_i = 268; \ \sum_{i=1}^{24} Y_i^2 = 6276; \ MS_R = 788.$$

On a aussi l'équation de prédiction suivante : $\hat{Y} = -50{,}681 + 0{,}850X_1 + 0{,}822X_2 - 1{,}231X_3$.

a) Au seuil 5%, peut-on dire qu'il y a au moins une variable X_i significative ? ***Oui, modèle globalement significatif car $F_0 = 17.41 > F_{3,20}(0.05)=3.10$***

b) Supposons que le modèle est satisfaisant, utiliser une loi d'approximation pour calculer le pourcentage d'observations (parmi les 24 considérées) dont la valeur absolue du résidu ($|e_i|$) dépasse $\sqrt{MS_E}$. ***Avec la loi normale : on trouve $\mathbf{2(1-\Phi(1)) = 0.31}$***

Les écarts-types des $\hat{\beta}_i$(estimateurs des β_i) sont donnés dans le tableau suivant :

i	0	1	2	3
$s(\hat{\beta}_i)$	14,811	0,126	0,356	0,385

c) Donner un intervalle de confiance de niveau 95% pour le paramètre β_3. Conclure.
[-2.034, -0.428]

d) Calculer la diminution attendue sur la variable Y si la valeur de X_3 passe de 100 à 150 (les valeurs de X_1 et X_2 restent inchangées). ***-61.55***

e) Sans effectuer aucun calcul, expliquer la procédure à suivre afin d'estimer les paramètres β_i, $i = 0, 1, 2, 3$, si le modèle à ajuster était de la forme :

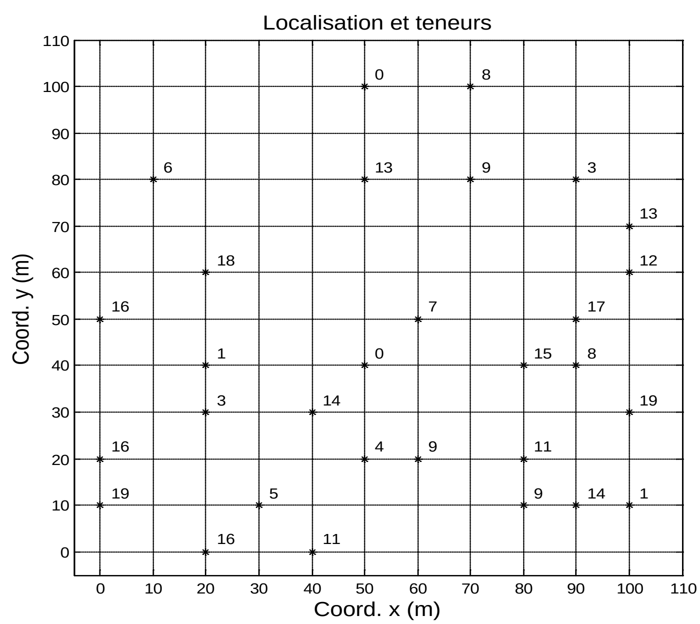
$$Y = \varepsilon e^{\beta_0 + \frac{\beta_1}{X_1} + \beta_2 X_2 + \beta_3 X_3}.$$

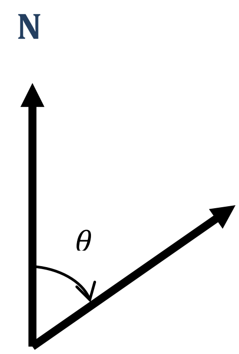
Poser Y' = lnY et X'=1/X ensuite chercher le vecteur B par calcul matriciel.

Réponse :

QUESTION 5 (8 points)

Calculer le variogramme expérimental des données suivantes dans la direction (azimut) $\theta = 63{,}43^{\text{o}}$ pour la distance h = $10\sqrt{5}$ m. Note : arctan(2) = 63,43 $^{\text{o}}$.





Réponse :

Il y a 4 paires. Résultat : 29.75.

<u>QUESTION 6</u> (16 points)

Un gisement de Zn présente un variogramme sphérique (2D) avec effet de pépite de 4 $\%^2$, C=20$\%^2$ et des portées suivant les différentes directions décrites par un modèle d'anisotropie géométrique avec les axes principaux en « x » et « y » et a_x=10m et a_y=20m. On désire estimer la teneur au point X_0 (situé en (0,0)) en se servant des teneurs mesurées aux points X_1 à X_3. Les coordonnées des points X_1 à X_3 sont respectivement (-10,0), (0,20) et (5,22). On a observé en ces points les teneurs respectivement Z_1=2%, Z_2=3.3% et Z_3=3%.

a) Montrer que le système de krigeage ordinaire, sous forme matricielle, permettant de calculer les poids des teneurs Z_i est le suivant :

$$\begin{bmatrix} 24 & 0 & 0 & 1 \\ 0 & 24 & 6.029 & 1 \\ 0 & 6.029 & 24 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

b) La résolution du système ci-haut donne μ = - 8,928 et λ_1 = 0,372. Calculer la valeur de la teneur estimée au point X_0. Que vaut la variance de krigeage ?
$\lambda_2 = \lambda_3 = 0.314$, Z_0(estimée) = 2.722% et variance de krigeage = 32.928$\%^2$

c) Observe-t-on un effet d'écran avec les résultats obtenus à la question b) ? Justifier.

d) Si au lieu d'avoir un modèle sphérique, on avait plutôt un modèle à effet de pépite pur, comment ceci affecterait-il les poids de krigeage et la variance de krigeage ?
$\lambda_1 = \lambda_2 = \lambda_3 = 1/3$ et variance de krigeage = 32$\%^2$

<u>Réponse :</u>

QUESTION 7 (12 points)

Afin d'établir l'impact de la précipitation maximale des dernières 48 heures (X en mm) sur le débit d'une rivière (Y en m^3/s), on prélève des données de la forme (X_i, Y_i) sur 19 rivières et on ajuste le modèle linéaire suivant :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

où β_0 et β_1 sont des paramètres et ε est une erreur aléatoire.

On dispose des sommes suivantes :

$$\sum_{i=1}^{19}(X_i - \bar{X})^2 = 488{,}33;\ \sum_{i=1}^{19}(X_i - \bar{X})(Y_i - \bar{Y}) = 1395.86;\ \sum_{i=1}^{19}(Y_i - \widehat{Y_i})^2 = 1260.$$

a) Au seuil 5%, la variable X a-t-elle une influence significative sur la variable Y ?

L'ajout de la variable T (surface du bassin hydrographique de la rivière, en km^2) au modèle précédent permet une amélioration du coefficient de détermination R^2. On passe ainsi de $R_1^2 = 0{,}76$ (modèle réduit) à $R_2^2 = 0{,}86$ (modèle complet). Test individuel : ***$|T0|$= 15.93 > $t_{17}(0.025)$ = 2.1, donc X significative.***

b) Que vaut SCE (SS_E) du modèle complet ? ***735***

Réponse :