Correction

	□0	□0	□0	□0	□0	□0	□0
	□1	□1	□1	□1	□1	□1	□1
	□2	□2	□2	□2	□2	□2	□2
	□3	□3	□3	□3	□3	□3	□3
	□4	□4	□4	□4	□4	□4	□4
	□5	□5	□5	□5	□5	□5	□5
	□6	□6	□6	□6	□6	□6	□6
	□7	□7	□7	□7	□7	□7	□7
□1	□8	□8	□8	□8	□8	□8	□8
Numéro étudiant : □2	□9	□9	□9	□9	□9	□9	□9

QCM Modèle linéaire

Durée : 15 minutes. Documents et calculatrice interdits
Téléphone, montre connectées et autres objets connectés : interdits
Une réponse correcte par question

Question 1 Pour introduire une covariable catégorielle à K modalités dans un modèle linéaire, combien de variables binaires doit-on introduire dans le modèle ?

■ $K-1$ □ $K+1$ □ K

Question 2 Un R^2 loin de 1 (par exemple, 0.2 ou 0.4) indique que le modèle de régression n'est pas utilisable.

■ Faux □ Vrai

Question 3 Sous l'hypothèse gaussienne, le modèle linéaire $Y = \beta_0 + \beta_1 X + \varepsilon$ s'écrit aussi

□ $Y \sim \mathcal{N}(\beta_0 + \beta_1 x;\ \sigma^2)$ □ $\left[Y \middle| X = x\right] \sim \mathcal{N}(\beta_0,\ \sigma^2)$
■ $\left[Y \middle| X = x\right] \sim \mathcal{N}(\beta_0 + \beta_1 x;\ \sigma^2)$ □ $Y \sim \mathcal{N}(\beta_0,\ \sigma^2)$

Question 4 Pour choisir/sélectionner des covariables, on peut utiliser

□ ridge □ des tests de nulllité des effets β_j ■ lasso

Question 5 On se place dans le modèle linéaire sans covariable $Y = \beta_0 + \varepsilon$, où la variance de ε est σ^2. On estime β_0 par moindre carré avec $\widehat{\beta}_0$ sur un échantillon de taille n. On veut également prédire un nouvel individu Y^*. On a

□ $\mathbb{E}\left[(\widehat{\beta}_0 - \beta_0)^2\right] = \sigma^2(1 + \frac{1}{n})$ et $\mathbb{E}\left[(\widehat{\beta}_0 - Y^*)^2\right] = \sigma^2$
□ $\mathbb{E}\left[(\widehat{\beta}_0 - \beta_0)^2\right] = \sigma^2/n$ et $\mathbb{E}\left[(\widehat{\beta}_0 - Y^*)^2\right] = \sigma^2$
■ $\mathbb{E}\left[(\widehat{\beta}_0 - \beta_0)^2\right] = \sigma^2/n$ et $\mathbb{E}\left[(\widehat{\beta}_0 - Y^*)^2\right] = \sigma^2(1 + \frac{1}{n})$
□ $\mathbb{E}\left[(\widehat{\beta}_0 - \beta_0)^2\right] = \sigma^2$ et $\mathbb{E}\left[(\widehat{\beta}_0 - Y^*)^2\right] = \sigma^2/n$

Question 6 Un effet β_j en facteur d'une covariable binaire $X_j \in \{0,1\}$ s'interprète comme une différence d'espérance entre des sous-population

□ Faux ■ Vrai

Question 7 Lorsque l'on introduit une covariables catégorielle Z à K modalités dans le modèle linéaire à côté d'une première covariable numérique X_1, on obtient :

□ des effets différents de X_1 pour chaque modalité de Z
■ des ordonnées à l'origine différentes pour chaque modalité de Z □ ni l'un, ni l'autre

Question 8 Pour utiliser l'hypothèse gaussienne, il faut vérifier que

☐ Y est gaussienne ☐ les covariables $X_1, \ldots, X_p$ sont gaussiennes
☐ R^2 n'est pas trop proche de 0 ■ aucun des trois précédents, mais autre chose

Question 9 On note $\mathbf{X}$ la matrice de design et $\mathbf{Y}$ le vecteur colonne des réponses observées. L'estimateur des moindres carrés $\widehat{\beta}$ dans le modèle linéaire est donné par

☐ $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{Y}$ ☐ $(\mathbf{X}'\mathbf{Y})^{-1}\mathbf{X}\mathbf{Y}$ ■ $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ☐ $(\mathbf{X}'\mathbf{X})\mathbf{X}'\mathbf{Y}$

Question 10 Avant d'utiliser le critère pénalisé Ridge, on doit

☐ vérifier l'hypothèse gaussienne ■ centrer-réduire les covariables

Question 11 Dans le modèle linéaire simple, le $i^{\text{ème}}$ résidu est

☐ $e_i = y_i - \beta_0 - \beta_1 x_i$ ■ $e_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$ ☐ $\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$
☐ $\varepsilon_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$

Question 12 Si on a obtenu $R^2 \approx 0.22$ sur un jeu de données, l'hypothèse de linéaire a peu de chance d'être vraie

☐ Vrai ■ Faux

Question 13 Comme graphique de diagnostic après l'inférence dans le modèle linéaire simple, on représente

☐ Les résidus e_i en fonction des valeurs observée de la covariable x_i
■ Les résidus e_i en fonction des réponses prédites $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$
☐ Les résidus e_i en fonction des réponses observées y_i

Question 14 Dans la régression multiple, l'effet β_j en facteur d'une covariable X_j est toujours du même signe que la covariance entre X_j et Y.

■ Faux ☐ Vrai

Question 15 En cas de sur-apprentissage,

☐ R^2 a tendance à être petit ☐ l'estimation $\widehat{\beta}$ est une très bonne estimation de β
■ l'estimation $\widehat{\sigma}^2$ sous-estime σ^2 ☐ aucun des trois