Modèle linéaire - QCM 1

STID2 - IUT VANNES

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5	6	7	8	9	10

1. Un modèle de régression simple permet :

A. d'expliquer le lien entre deux variables qualitatives.

B. d'expliquer le lien entre deux variables quantitatives.

C. d'expliquer le lien entre une variable qualitative et une variable quantitative.

2. Si la corrélation entre deux variables x et y vaut 0, alors :

A. la pente du modèle de régression simple vaut 1.

B. la pente du modèle de régression simple vaut 0.

C. la pente du modèle de régression simple vaut -1.

3. Dans une régression simple, l'estimateur aux moindres carrés $\hat{\beta}_1$ de la pente est (en notant y la variable à expliquer et x la variable explicative) :

A. $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

B. $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$.

C. $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}$.

D. $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$.

4. Dans une régression simple, l'estimateur aux moindres carrés $\hat{\beta}_0$ de l'ordonnée à l'origine est (en notant y la variable à expliquer et x la variable explicative) :

A. $\hat{\beta}_0 = \bar{y} + \hat{\beta}_1\bar{x}$.

B. $\hat{\beta}_0 = \bar{y} - \beta_1\bar{x}$.

C. $\hat{\beta}_0 = \bar{y} + \beta_1\bar{x}$.

D. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$.

5. Les estimateurs aux moindres carrés $\hat{\beta}_0$ (pour l'ordonnée à l'origine) et $\hat{\beta}_1$ (pour la pente) de la régression simple sont sans biais car :

A. $\mathbb{E}(\hat{\beta}_0) = 0$ et $\mathbb{E}(\hat{\beta}_1) = 0$.

B. $V(\hat{\beta}_0) = 0$ et $V(\hat{\beta}_1) = 0$.

C. $\mathbb{E}(\hat{\beta}_0) = \beta_0$ et $\mathbb{E}(\hat{\beta}_1) = \beta_1$.

D. $V(\hat{\beta}_0) = \beta_0$ et $V(\hat{\beta}_1) = \beta_1$.

6. Dans une régression simple expliquant une variable y par une variable x, le résidu $\hat{e}_i$ pour un individu i correspond graphiquement :

A. à l'écart vertical entre le point (x_i, y_i) et la droite de régression ajustée.

B. à l'écart horizontal entre le point (x_i, y_i) et la droite de régression ajustée.

C. à la pente de la droite.

D. à l'ordonnée à l'origine de la droite.

7. Dans un modèle de régression simple, la variance des erreurs peut être estimée par la formule :

A. $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{e}_i$

B. $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{e}_i^2$

C. $\hat{\sigma}^2 = (n-2) \sum_{i=1}^{n} \hat{e}_i^2$

8. Comment s'écrit le modèle ajusté pour une régression linéaire simple expliquant y par x :

A. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}$

B. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

C. $\hat{y} = \beta_0 + \beta_1 x$

9. Que vaut la somme des résidus d'une régression ?

A. 1

B. σ

C. σ^2

D. 0

10. Quelle relation est vraie ? (en notant ρ_{xy} la corrélation entre x et y, s_x et s_y les écarts-types de x et y)

A. $\hat{\beta}_1 = \rho_{xy} \frac{s_y^2}{s_x^2}$

B. $\hat{\beta}_1 = \rho_{xy} \frac{s_x^2}{s_y^2}$

C. $\hat{\beta}_1 = \rho_{xy} \frac{s_y}{s_x}$

D. $\hat{\beta}_1 = \rho_{xy} \frac{s_x}{s_y}$

Modèle linéaire - QCM 2 (sujet A)

STID2 - IUT VANNES

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5	6	7	8	9	10

1. Dans une régression simple, le coefficient de détermination R^2 est égal (en notant y la variable à expliquer et x la variable explicative) :

A. au carré du coefficient de corrélation empirique entre les x_i et les y_i.

B. à la racine carrée du coefficient empirique entre les x_i et les y_i.

C. au coefficient de corrélation empirique entre les x_i et les y_i.

2. L'intervalle de confiance à 95% d'un coefficient d'un modèle de régression ne contient pas 0. Laquelle des affirmations suivantes est correcte ?

A. Au niveau 5%, le coefficient est significativement différent de 0.

B. Au niveau 5%, le coefficient n'est pas significativement différent de 0.

C. L'intervalle de confiance ne permet pas de conclure sur la significativité du coefficient.

3. Dans une régression linéaire simple, quelle est la statistique de test pour le test de Student de significativité du paramètre de pente β_1 ?

A. $T = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}$.

B. $T = \frac{\hat{\beta}_1}{\hat{\sigma}^2_{\hat{\beta}_1}}$.

C. $T = \frac{\beta_1}{\sigma_{\hat{\beta}_1}}$.

D. $T = \frac{\beta_1}{\sigma^2_{\hat{\beta}_1}}$.

4. Une régression simple a été effectuée à partir de n individus, avec $n = 32$. Le quantile de niveau $1 - \frac{\alpha}{2}$ de la loi de Student $St(n-2)$ vaut $1,7$ pour $\alpha = 0,05$. La statistique de test du test de Student de significativité du paramètre de pente β_1 renvoie la valeur $-1,3$. Dans ce cas :

A. Au niveau α, on ne rejette pas $H_0 : \beta_1 = 0$.

B. Au niveau α, on rejette $H_0 : \beta_1 = 0$.

5. Quel test est le plus important quand on teste la significativité des coefficients d'une régression simple ? :

A. Le test de Student permettant de tester la nullité de la pente.

B. Le test de Student permettant de tester la nullité de l'ordonnée à l'origine.

C. Les deux tests ont la même importance.

6. On met en place une régression simple expliquant le niveau de pollution maximum de la journée par la température maximum de la journée. Grâce au test de Student, nous avons réussi à conclure que le paramètre de pente β_1 est significativement différent de 0. On peut alors conclure :

A. La température a bien un effet sur le niveau de pollution.

B. La température n'a pas d'effet sur le niveau de pollution.

7. À quoi sert un modèle de régression multiple ?

A. Expliquer une variable quantitative par plusieurs variables qualitatives.

B. Expliquer une variable qualitative par plusieurs variables quantitatives.

C. Expliquer une variable quantitative par plusieurs variables quantitatives.

8. Dans un modèle de régression linéaire multiple, on note X la matrice des observations des variables explicatives et Y le vecteur des observations de la variable à expliquer. L'estimateur des moindres carrés $\hat{\beta}$ du coefficient de régression β vaut :

A. $\hat{\beta} = (XX^T)^{-1}X^TY$.

B. $\hat{\beta} = (X^TX)^{-1}X^TY$.

C. $\hat{\beta} = X^T(X^TX)^{-1}Y$.

D. $\hat{\beta} = X^T(XX^T)^{-1}Y$.

9. Si le modèle de régression multiple contient une constante β_0 :

A. La première colonne de la matrice X des variables explicatives ne contient que des 1.

B. La première colonne de la matrice X des variables explicatives ne contient que des 0.

C. La première colonne de la matrice X des variables explicatives contient des valeurs entre -1 et 1.

10. On donne la sortie R ci-dessous pour une régression simple :

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 134.3450    45.4737   2.954  0.00386 **
superficie    6.6570     0.6525  10.203  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 77.93 on 106 degrees of freedom
Multiple R-squared: 0.4955,     Adjusted R-squared: 0.4907
F-statistic: 104.1 on 1 and 106 DF,  p-value: < 2e-16
```

Que vaut $\hat{\sigma}^2_{\hat{\beta}_1}$ (variance estimée de l'estimateur de la pente $\hat{\beta}_1$) ?

A. 45.4737

B. 45.4737^2

C. 0.6525

D. 0.6525^2

Modèle linéaire - QCM 3

STID2 - IUT VANNES - a

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5	6	7	8	9	10

1. Une régression a été effectuée et on relève son coefficient de détermination R^2. On ajoute ensuite une variable explicative au modèle. Que peut-on dire sur le nouveau coefficient de détermination calculé ?

A. Il est plus faible.

B. Il est plus élevé.

C. Sa valeur n'a pas changé.

2. Le coefficient de détermination R^2 d'un modèle de régression ne contenant pas la constante β_0 est :

A. égal au R^2 du même modèle avec constante.

B. inférieur au R^2 du même modèle avec constante.

C. supérieur au R^2 du même modèle avec constante.

3. On met en place un modèle de régression simple pour expliquer le prix du forfait dans une station par le nombre de pistes dans cette station. Le code R est le suivant :

```
reg_ski = lm (Forfait ~ NbPistes, data=Ski)
summary(reg_ski)
```

Que contient `reg_ski$fitted.values` ?

A. les valeurs de la variable à expliquer : $\text{Forfait}_i \quad \forall i = 1, \ldots, n$

B. les valeurs ajustées pour la variable à expliquer : $\hat{\text{Forfait}}_i \quad \forall i = 1, \ldots, n$

C. les valeurs des résidus $\hat{e}_i \quad \forall i = 1, \ldots, n$

4. On donne la sortie de l'ajustement de ce modèle :

```
Call:
lm(formula = Forfait ~ NbPistes, data = Ski)

Residuals:
    Min      1Q  Median      3Q     Max
-55.927 -14.775  -2.104  12.728  92.277

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 88.45930    4.59585   19.25   <2e-16 ***
NbPistes     0.87338    0.07615   11.47   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.79 on 96 degrees of freedom
Multiple R-squared:  0.5781,Adjusted R-squared:  0.5737
F-statistic: 131.6 on 1 and 96 DF,  p-value: < 2.2e-16
```

Comment s'écrit le modèle ajusté ?

A. $\widehat{\text{Forfait}} = 88.46 + 0.87 * \text{NbPistes}$

B. $\widehat{\text{Forfait}} = 0.87 + 88.46 * \text{NbPistes}$

C. $\widehat{\text{Forfait}} = 88.46 + 0.87 * \text{NbPistes} + 24.79$

D. $\widehat{\text{Forfait}} = 0.87 + 88.46 * \text{NbPistes} + 24.79$

5. Dans un modèle de régression multiple, quelle affirmation est vraie ?

A. La variable à expliquer ne doit pas être trop corrélée aux variables explicatives.

B. Les variables explicatives ne doivent pas être trop corrélées entre elles.

C. Les variables explicatives doivent être très corrélées entre elles.

6. Combien de modèles de régression différents peut-on construire à partir d'un ensemble de 3 variables explicatives ?

A. 3

B. 6

C. 7

D. 9

7. On veut prédire le prix du forfait pour une station située à une altitude de 2000m et ayant 30 pistes. On utilise la prédiction issue du modèle de régression multiple nommé "reg_ski_mult", et on obtient le résultat ci-dessous :

```
> newdata = data.frame(NbPistes=30,AltPistes=2000)
> predict(reg_ski_mult,newdata,se.fit=TRUE)
$`fit`
       1
108.0494

$se.fit
[1] 2.619112

$df
[1] 95

$residual.scale
[1] 20.41051
```

Quelle information nous donne la valeur "se.fit" ?

A. Une estimation de l'écart-type d'erreur de prévision.

B. Une estimation de l'écart-type des erreurs du modèle.

C. La valeur prédite pour le prix du forfait.

8. Nous avons 3 variables x_1, x_2 et x_3 à disposition pour expliquer une variable y. Nous voulons choisir entre les deux modèles suivants : un modèle expliquant y par x_1 ; un modèle expliquant y par x_2 et x_3. Quelle stratégie peut-on employer ?

A. Utiliser le test de Fisher emboîté.

B. Utiliser le critère du R^2 ajusté.

C. Aucune de ces stratégies ne permet de conclure.

9. Nous avons 4 variables x_1, x_2, x_3 et x_4 à disposition pour expliquer une variable y. Nous mettons en place un test de Fisher emboîté pour choisir entre les deux modèles suivants : un modèle expliquant y par toutes les variables disponibles ; un modèle expliquant y par x_1 et x_2. Comment s'écrit l'hypothèse H_0 du test ?

A. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

B. $H_0 : \beta_1 = \beta_2 = 0$

C. $H_0 : \beta_3 = \beta_4 = 0$

10. On note X la matrice des variables explicatives d'un modèle de régression multiple à 2 variables. On donne le contenu de la matrice $(X'X)^{-1}$:

$$(X'X)^{-1} = \begin{pmatrix} 0.693 & -0.005 & -0.462 \\ -0.005 & 0.001 & -0.014 \\ -0.462 & -0.014 & 20.716 \end{pmatrix}$$

et la valeur de la variance estimée des erreurs du modèle : $\hat{\sigma}^2 = 10$.
On note $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ l'estimateur des coefficients de la régression.

Que vaut la variance estimée $\hat{V}(\hat{\beta}_1)$ (notée aussi $\hat{\sigma}^2_{\hat{\beta}_1}$) de l'estimateur $\hat{\beta}_1$?

A. 0.01

B. 0.001

C. 0.0001

D. 0.693

E. 6.93

Modèle linéaire - QCM 4

STID2 - IUT VANNES - a

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5	6	7	8	9	10

1. Nous avons 3 variables x_1, x_2 et x_3 à disposition pour expliquer une variable y. Nous voulons choisir entre les deux modèles suivants : un modèle expliquant y par x_1 ; un modèle expliquant y par x_2 et x_3. Quelle stratégie peut-on employer ?

A. Utiliser le test de Fisher emboîté.

B. Utiliser le critère du R^2 ajusté.

C. Aucune de ces stratégies ne permet de conclure.

2. Nous avons 4 variables x_1, x_2, x_3 et x_4 à disposition pour expliquer une variable y. Nous mettons en place un test de Fisher emboîté pour choisir entre les deux modèles suivants : un modèle expliquant y par toutes les variables disponibles ; un modèle expliquant y par x_1 et x_2. Comment s'écrit l'hypothèse H_0 du test ?

A. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

B. $H_0 : \beta_1 = \beta_2 = 0$

C. $H_0 : \beta_3 = \beta_4 = 0$

3. Nous avons 4 variables x_1, x_2, x_3 et x_4 à disposition pour expliquer une variable y. Nous voulons choisir entre les deux modèles suivants : un modèle expliquant y par x_1, x_2 et x_3 ; un modèle expliquant y par x_2. Quelle stratégie peut-on employer ?

A. Utiliser le test de Fisher emboîté.

B. Utiliser le critère du R^2 ajusté.

C. Aucune de ces stratégies ne permet de conclure.

D. Les deux stratégies permettent de conclure.

4. Dans un modèle de régression simple expliquant une variable y par une variable explicative x, on souhaite tester la significativité du coefficient de pente β_1. On peut utiliser :

A. Uniquement le test de Fisher global.

B. Uniquement le test de Student.

C. Les deux tests car ils donneront la même conclusion.

5. Dans un modèle de régression multiple contenant une constante β_0 et 3 variables explicatives x_1, x_2 et x_3, on souhaite tester la significativité du coefficient β_2 associé à x_2. On peut utiliser :

A. Uniquement le test de Fisher global.

B. Uniquement le test de Student.

C. Les deux tests car ils donneront la même conclusion.

6. À quoi sert le VIF ("Variance Inflation Factor") ?

A. À vérifier que les variables explicatives d'un modèle de régression multiple ne sont pas trop corrélées entre elles.

B. À étudier la présence d'individus atypiques.

C. À étudier la présence d'individus influents.

D. À vérifier que la variance de l'estimateur des moindres carrés n'est pas trop élevée.

7. Pour vérifier graphiquement que les résidus d'un modèle linéaire suivent une loi gaussienne, on peut regarder :

A. L'histogramme des résidus.

B. Le boxplot des résidus.

C. Les deux graphiques permettent de répondre à la question.

8. Afin de vérifier que l'hypothèse d'homoscédasticité des erreurs est vérifiée, on regarde :

A. Le nuage de points des résidus studentisés en fonction des valeurs ajustées de la variable à expliquer.

B. L'histogramme des résidus studentisés.

C. Le qq-plot (ou "droite de Henry") des résidus studentisés.

9. Après avoir mis en place un modèle de régression ou une analyse de la variance, on a détecté un individu atypique. Alors :

a. On supprime cet individu et on relance l'estimation du modèle.

b. On essaye de comprendre/expliquer pourquoi cet individu a un résidu élevé.

c. On modifie les données correspondant à cet individu afin qu'il ne soit plus atypique.

10. On a ajusté un modèle de régression multiple expliquant le prix des forfaits par le nombre de remontées dans la station et l'altitude du sommet des pistes. La sortie donnée par le logiciel R est la suivante :

```
lm(formula = Forfait ~ NbRemontees + AltPistes, data = Ski)

Residuals:
    Min      1Q  Median      3Q     Max
-76.416  -8.571  -0.695   9.226  48.679

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.494035   9.413797   2.496   0.0143 *
NbRemontees  0.858483   0.104148   8.243 9.85e-13 ***
AltPistes    0.033132   0.004331   7.650 1.72e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.24 on 94 degrees of freedom
Multiple R-squared: 0.7226,     Adjusted R-squared: 0.7167
F-statistic: 122.4 on 2 and 94 DF,  p-value: < 2.2e-16
```

On peut conclure à partir de cette sortie que les deux paramètres $\beta_{NbRemontees}$ et $\beta_{AltPistes}$ sont significativement différents de 0. Quelles informations dans la sortie nous ont permis d'aboutir à cette conclusion :

A. Les deux tests de Student pour $\beta_{NbRemontees}$ et $\beta_{AltPistes}$.

B. Le test de Fisher global.

C. Les deux : les tests de Student et le test de Fisher global permettent de donner cette conclusion.

Modèle linéaire - QCM 5

STID2 - IUT VANNES - a

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5

1. Après avoir estimé un modèle de régression, on a identifié un individu ayant une distance de Cook élevée et on sait déjà que cet individu n'est pas atypique. Alors :

A. On ne peut rien conclure sur l'influence de cet individu.

B. Cet individu est influent.

C. Cet individu n'est pas influent.

2. Pour choisir les variables à inclure dans un modèle de régression multiple, on met en place une procédure de sélection de variables de type "backward" basée sur des tests de Student de significativité. On part du modèle complet contenant toutes les variables disponibles. Que fait-on à chaque étape intermédiaire de la procédure ?

A. On ajoute la variable ayant la p-value la plus petite.

B. On ajoute la variable ayant la p-value la plus grande.

C. On retire la variable ayant la p-value la plus petite.

D. On retire la variable ayant la p-value la plus grande.

3. Combien de modèles de régression différents peut on construire à partir d'un ensemble de p variables explicatives ?

A. 2^{p-1}.

B. $2^p - 1$.

C. 2^{p+1}.

D. $2^p + 1$.

E. 2^p.

4. Pour choisir les variables à inclure dans un modèle de régression multiple, on met en place une procédure de sélection de variables de type "forward" basée sur le critère BIC. Quand arrête-t-on la procédure ?

A. Lorsque le critère BIC ne diminue plus.

B. Lorsque le critère BIC n'augmente plus.

C. Lorsque tous les tests de Student sont significatifs.

5. Que constate-t-on souvent en pratique ?

A. Une procédure de choix de modèle basée sur le BIC conduit souvent à des modèles plus "gros" qu'une procédure basée sur le R^2 ajusté.

B. Une procédure de choix de modèle basée sur le BIC conduit souvent à des modèles plus "petits" qu'une procédure basée sur le R^2 ajusté.

C. Les deux procédures conduisent toujours à des modèles de même taille.

Modèle linéaire - QCM 6

STID2 - IUT VANNES - a

Année 2021/2022

NOM :

Prénom :

Groupe :

Réponses aux questions :

1	2	3	4	5	6

1. Dans le modèle d'analyse de la variance : $y_{ik} = \mu_i + e_{ik}$, que représente y_{ik} ?

A. La valeur prise par la variable y pour l'individu k du groupe i.

B. La valeur moyenne de y dans le groupe i.

C. La valeur moyenne de y dans le groupe k.

D. La valeur prise par la variable y pour l'individu i du groupe k.

2. Pour le jeu de données suivant :

Individu	O3	vent
1	64	E
2	90	N
3	79	E
4	81	S
5	88	O

comment va s'écrire la matrice X des variables explicatives construite pour l'analyse de variance :

A.
$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

B.
$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

C.
$$X = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

3. Dans le modèle d'analyse de la variance : $y_{ik} = \mu_i + e_{ik}$, que vaut l'estimation $\hat{\mu}_i$ du paramètre μ_i ?

A. $\bar{y}_i$

B. $\hat{y}_i$

C. $\bar{y}_i - \bar{y}_1$

D. $\hat{y}_i - \hat{y}_1$

4. Dans le modèle d'analyse de la variance : $y_{ik} = \mu_i + e_{ik}$, que vaut la variance $\hat{\sigma}^2$?

A. $\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\sum_{k=1}^{n_i}(y_{ik}-\bar{y}_i)^2}{n-1}$

B. $\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\sum_{k=1}^{n_i}(\hat{y}_i-\bar{y}_i)^2}{n-1}$

C. $\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\sum_{k=1}^{n_i}(y_{ik}-\bar{y}_i)^2}{n-I}$

D. $\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\sum_{k=1}^{n_i}(\hat{y}_i-\bar{y}_i)^2}{n-I}$

5. À partir de la sortie R suivante, quelle valeur trouve-t-on pour la valeur moyenne de "maxO3" lorsque la modalité de la variable "vent" est "Ouest" ?

```
Call:
lm(formula = maxO3 ~ vent, data = ozone)

Residuals:
    Min      1Q  Median      3Q     Max
-60.600 -16.807  -7.365  11.478  81.300

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  105.600      8.639  12.223   <2e-16 ***
ventNord     -19.471      9.935  -1.960   0.0526 .
ventOuest    -20.900      9.464  -2.208   0.0293 *
ventSud       -3.076     10.496  -0.293   0.7700
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.32 on 108 degrees of freedom
  (10 observations deleted due to missingness)
Multiple R-squared: 0.08602,    Adjusted R-squared: 0.06063
F-statistic: 3.388 on 3 and 108 DF,  p-value: 0.02074
```

A. $105.6 - 20.9$

B. $105.6 + 20.9$

C. $105.6 - 19.471 - 20.9$

D. $105.6 + 19.471 + 20.9$

6. Pourquoi le chapitre 5 s'appelle "analyse de la variance" ?

A. Car l'objectif est de comparer des variances dans plusieurs groupes.

B. Car la statistique de test du test de Fisher utilisé est un rapport de variances.

C. Car l'analyse repose uniquement sur la variance des erreurs du modèle.