

3.6 Test F de l'importance globale de la régression

Un test d'hypothèse important en régression consiste à tester si au moins une des variables exogènes explique une partie significative de la variabilité dans les Y_i . Ceci revient donc à tester si les données permettent de réfuter l'hypothèse nulle H_0 à l'effet que les variables exogènes n'expliquent rien. Mathématiquement, une variable exogène n'explique en rien la valeur de Y_i si le coefficient de régression correspondant prend la valeur 0. On veut donc confronter les hypothèses

$$H_0 : \quad \beta_1 = \cdots = \beta_{p'} = 0 \quad (3.14)$$

$$H_1 : \quad \text{au moins un des coefficients n'est pas nul.}$$

Sous H_0 , le modèle de régression ne devrait pas expliquer la variabilité dans les Y_i et donc le ratio SS_{Reg}/SS_E devrait prendre une petite valeur. Par contre sous H_1 , le modèle de régression devrait expliquer une partie de la variabilité des Y_i et donc le ratio SS_{Reg}/SS_E devrait prendre une grande valeur. Afin de savoir si la valeur du ratio est “petite” ou “grande,” on standardise le ratio pour obtenir la statistique F du tableau d'ANOVA :

$$F = \frac{SS_{Reg}/p'}{SS_E/(n-p)} = \frac{SS_E/p'}{s^2}.$$

Sous H_0 , cette statistique suit une loi F avec p' degrés de liberté au numérateur et $n-p$ degrés de liberté au dénominateur. On rejette donc H_0 au seuil α (on admet donc que les données suggèrent que le modèle n'est pas complètement inutile ou encore qu'il existe une relation entre la variable endogène et au moins une des variables exogènes) lorsque la statistique F est supérieure ou égale au quantile $F_{\alpha; p', n-p}$.

Exemple 3.4 (Consommation d'essence) *La statistique F obtenue pour le test d'importance globale de la régression est de 22.70 sur 4 et 43 degrés de liberté, respectivement. Ceci*

équivalent à un seuil observé de $P(F_{4,43} \geq 22.70) \approx 0$ et donc nous rejetons l'hypothèse nulle. En d'autres termes, il est clair qu'au moins un des coefficients de régression n'est pas nul.

3.7 Test de réduction du modèle

Il sera plutôt rare en pratique que le test F global ne rejette pas l'hypothèse H_0 donnée par (3.14) que le modèle de régression est totalement inutile. Cependant, il se peut qu'on veuille tester si le modèle peut être réduit. Cela revient à tester si un sous-modèle plus simple explique une partie suffisamment grande de la variabilité dans les Y_i pour qu'il ne soit pas nécessaire d'utiliser le modèle plus complexe.

Le principe de la somme de carrés résiduels additionnelle permet de tester cette hypothèse de façon formelle. L'idée est simple : si les termes qui sont exclus du modèle plus simple expliquent une partie importante de la variabilité dans les Y_i , alors la variabilité due à la fluctuation aléatoire (SS_E) apparaîtra beaucoup plus importante dans le modèle simple que dans le modèle complet. Il s'agit donc de mesurer si la différence entre les sommes de carrés résiduelles des deux modèles est petite ou grande.

Mathématiquement, supposons que l'on a le modèle de régression multiple

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p'} x_{ip'} + \varepsilon_i$$

et que l'on veut tester si un modèle avec seulement $k < p'$ des variables exogènes suffirait à expliquer la variabilité dans les Y_i , ou autrement dit que $p' - k$ des variables exogènes sont superflues. Par simplicité, supposons que les k variables exogènes en question sont x_1, \dots, x_k . On veut alors confronter les hypothèses

$$H_0 : Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (3.15)$$

$$H_1 : Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \beta_{k+1} x_{i,k+1} + \cdots + \beta_{p'} x_{ip'} + \varepsilon_i.$$

Nous appellerons “modèle réduit” le modèle donné par H_0 et “modèle complet” le modèle donné en H_1 . Soit $SS_E^{H_0}$ la somme des carrés résiduelle associée au modèle réduit et $SS_E^{H_1}$ la somme des carrés résiduelle associée au modèle complet. Alors il est toujours vrai que $SS_E^{H_1} \leq SS_E^{H_0}$. Cependant, sous H_0 , la différence entre $SS_E^{H_0}$ et $SS_E^{H_1}$ sera petite, alors que cette même différence sera grande sous H_1 . Encore une fois, on standardise le ratio $(SS_E^{H_0} - SS_E^{H_1})/SS_E^{H_1}$ afin d’obtenir une distribution connue qui nous permettra d’identifier si une valeur est “petite” ou “grande.” On obtient la statistique de test suivante:

$$F = \frac{(SS_E^{H_0} - SS_E^{H_1})/\Delta_{dl}}{SS_E^{H_1}/(n-p)} = \frac{SS_E^{H_0} - SS_E^{H_1}}{\Delta_{dl}s_{H_1}^2}, \quad (3.16)$$

où Δ_{dl} est la différence entre les degrés de liberté de $SS_E^{H_0}$ et les degrés de liberté de $SS_E^{H_1}$.

Sous H_0 donnée par (3.15), la statistique F en (3.16) obéit à une loi de Fisher–Snedecor avec Δ_{dl} degrés de liberté au numérateur et $n-p$ degrés de liberté au dénominateur. On rejette donc H_0 au seuil α lorsque $F \geq F_{\alpha; \Delta_{dl}, n-p}$. Notez que $n-p$ fait référence aux degrés de liberté de $SS_E^{H_1}$ et que $s_{H_1}^2$ réfère au carré moyen résiduel sous H_1 . Comme il arrivera parfois que notre “modèle complet” sera un modèle qui n’inclut pas toutes les p' variables exogènes, il faudra ajuster la formule (3.16) en conséquence. C’est pourquoi il est probablement mieux de retenir la forme à droite de la seconde égalité en (3.16).

Il existe plusieurs moyens pour calculer Δ_{dl} :

1. $\Delta_{dl} = dl(SS_E^{H_0}) - dl(SS_E^{H_1})$;
2. $\Delta_{dl} = (\text{nombre de paramètres modèle } H_1) - (\text{nombre de paramètres modèle } H_0)$;
3. $\Delta_{dl} = \text{nombre de contraintes sur les paramètres du modèle complet pour arriver au modèle réduit.}$

Ainsi, dans le cas présenté en (3.15), on aurait

1. $dl(SS_E^{H_0}) = n - (k + 1) = n - k - 1$ et $dl(SS_E^{H_1}) = n - p$, d'où $\Delta_{dl} = n - k - 1 - n + p = p - k - 1 = p' - k$.
2. Paramètres sous H_1 : $p = p' + 1$, paramètres sous H_0 : $k + 1$, et donc $\Delta_{dl} = p' + 1 - k - 1 = p' - k$.
3. Les contraintes pour réduire le modèle en H_1 au modèle en H_0 sont $\beta_{k+1} = 0, \dots, \beta_{p'} = 0$, c'est-à-dire $p' - k$ égalités (contraintes) sur les paramètres du modèle complet afin d'obtenir le modèle réduit, donc $\Delta_{dl} = p' - k$.

La très vaste majorité des tests d'hypothèses que nous aurons à confronter dans ce cours pourront être exprimées sous la forme

$$H_0 : \text{modèle réduit}, \quad H_1 : \text{modèle complet}.$$

Par exemple le test F du tableau d'ANOVA est un test de cette forme, où le modèle réduit est tout simplement $Y_i = \beta_0 + \varepsilon_i$. Dans ce cas, on peut voir facilement que la statistique F donnée en (3.16) est égale à la statistique F de la table ANOVA (exercice). De plus, on voit aussi facilement que $\Delta_{dl} = p'$.

Remarque 3.1 *Sous le postulat de normalité des termes d'erreur, il est possible de démontrer que le test F du principe de la somme de carrés résiduels additionnelle est un test du rapport des vraisemblances.*

Exemple 3.5 (Consommation d'essence) *Un économiste bien connu postule que ni la longueur des routes fédérales ni la proportion de détenteurs de permis de conduire n'ont un impact sur la consommation d'essence. Testez cette hypothèse au seuil de 5%.*

Le modèle sous H_0 dans ce cas est donné par $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$. Le tableau d'ANOVA de ce modèle est

Source	d.l.	SC	CM	F
Régression	2	153 478	76 739	7.94
Résiduelle	45	434 889	9664	
Totale	47	588 366		

On obtient donc une statistique F pour le test :

$$\begin{aligned}
 F &= \frac{SS_E^{H_0} - SS_E^{H_1}}{\Delta_{dl} s^2} = \frac{434\,889 - 189\,050}{(45 - 43) \times 4397} \\
 &= 27.96.
 \end{aligned}$$

En consultant une table de la loi de Fisher-Snedecor, on voit que $F_{0.05; 2, 43} = 3.21$. Comme $27.96 > 3.21$, on rejette l'hypothèse émise par l'économiste.

3.8 Test d'une hypothèse linéaire générale sur β

Plusieurs tests en régression linéaire se veulent en fait une liste de contraintes linéaires sur les valeurs des paramètres. Soit \mathbf{C} une matrice de dimension $r \times p$ et \mathbf{d} un vecteur de dimension $r \times 1$. Alors une hypothèse linéaire générale est une hypothèse de la forme

$$H_0 : \mathbf{C}\beta = \mathbf{d}, \quad H_1 : \mathbf{C}\beta \neq \mathbf{d}. \quad (3.17)$$

Exemple 3.6 (Consommation d'essence) *Tester l'hypothèse qu'une hausse de la taxe de vente de 1% a exactement le même impact sur la consommation d'essence moyenne qu'une baisse du revenu de 500 \$.*

En termes mathématiques, une hausse de 1% de la taxe sur la consommation d'essence moyenne est donnée par

$$\begin{aligned}
 E(Y; x_1 + 1) - E(Y; x_1) &= \beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \beta_3 x_3 \\
 &\quad - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4) \\
 &= \beta_1.
 \end{aligned}$$

De façon similaire, on trouve que l'effet d'une baisse du revenu de 500 \$ (x_3 diminuée de 0.5) sur la consommation moyenne est $-0.5\beta_3$. On veut donc tester

$$H_0 : \beta_1 = -0.5\beta_3 \Leftrightarrow \beta_1 + 0.5\beta_3 = 0,$$

c'est-à-dire que $\mathbf{C} = (0, 1, 0, 0.5, 0)$ et $\mathbf{d} = 0$.

On peut réécrire le test d'hypothèses 3.17 sous la forme d'un modèle réduit et d'un modèle complet :

$$H_0 : \mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad H_1 : \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où l'équation en H_0 est obtenue en (i) résolvant le système d'équations $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ pour $\boldsymbol{\beta}$ et (ii) en substituant la solution trouvée en (i) dans le modèle $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Puisque le test d'une hypothèse linéaire générale peut être exprimé comme le test d'un modèle réduit par rapport à un modèle complet, on peut utiliser le principe de somme des carrés résiduelle additionnelle. Il est possible de démontrer (voir par exemple Sen & Srivastava 1990, pp. 44–45 et pp. 60–64) que

$$SS_E^{H_0} - SS_E^{H_1} = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \{\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top\}^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}).$$

Noter que $\sigma^2 \{\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top\}$ est la variance de $\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}$. Donc la différence des sommes de carrés résiduels peut être vue comme la longueur du vecteur qui sépare $\mathbf{C}\hat{\boldsymbol{\beta}}$ et \mathbf{d} , standardisée par sa variance.

Si la distance entre $\mathbf{C}\hat{\boldsymbol{\beta}}$ et \mathbf{d} est petite, alors H_0 est une hypothèse raisonnable, $SS_E^{H_0} - SS_E^{H_1}$ prend une petite valeur et on ne rejette donc pas H_0 . Si H_0 n'est pas raisonnable, alors $\mathbf{C}\hat{\boldsymbol{\beta}}$ sera loin de \mathbf{d} , $SS_E^{H_0} - SS_E^{H_1}$ prendra une grande valeur et nous rejetterons H_0 . On peut maintenant terminer la construction de la statistique F en utilisant l'équation (3.16).

On a donc

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^\top \{\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top\}^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})/r}{SS_E^{H_1}/(n-p)} \quad (3.18)$$

ou encore

$$F = \frac{(C\hat{\beta} - d)^\top \{C(X^\top X)^{-1}C^\top\}^{-1}(C\hat{\beta} - d)}{rs_{H_1}^2}, \quad (3.19)$$

où r est le nombre de lignes de la matrice C . Le fait que $\Delta_{dl} = r$ dans ce cas est facile à comprendre si on utilise le truc 3, soit le nombre de contraintes sur les paramètres du modèle complet pour obtenir le modèle réduit. Sous l'hypothèse H_0 , la statistique F suit une loi de Fisher:

$$F \sim \text{Fisher}(r, n - p)$$

Remarque 3.2 Dans l'aide de SAS, on utilise la notation $L\beta = c$ plutôt que la notation $C\beta = d$ qui est utilisée par plusieurs livres. L'aide de SAS utilise aussi b plutôt que $\hat{\beta}$ pour dénoter l'estimateur de β .

Exemple 3.7 (Consommation d'essence) En reprenant le contexte présenté au début de la section, on veut donc tester

$$H_0 : \beta_1 = -0.5\beta_3 \Leftrightarrow \beta_1 + 0.5\beta_3 = 0,$$

c'est-à-dire que $C = (0, 1, 0, 0.5, 0)$ et $d = 0$. On obtient

$$\{C(X^\top X)^{-1}C^\top\}^{-1} = 20.32 \quad \text{et} \quad C\hat{\beta} - d = -68.1.$$

Par conséquent,

$$F = \frac{(-68.1) \times (20.32) \times (68.1)}{(1) \times (4397)} = 21.4.$$

Ceci équivaut à un seuil de $P(F_{1,43} \geq 21.4) < 0.0001$, ce qui conduit au rejet de H_0 .

Noter que nous aurions aussi pu faire le test de $H_0 : Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$ de la section 3.3 avec la méthode du test d'une hypothèse linéaire généralisée, puisque dans ce cas

H_0 peut se récrire comme $H_0 : \beta_2 = 0, \beta_4 = 0$. Il s'agit donc d'un test d'hypothèse linéaire généralisée avec $r = 2$:

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{et} \quad \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

3.9 Géométrie de l'analyse de variance et des tests d'hypothèses

Nous avons vu plus tôt comment les sommes de carrés de l'analyse de la variance peuvent être vues comme des carrés de la norme de vecteurs. Ceci est également vrai en ce qui concerne le test F du principe de somme des carrés résiduelle additionnelle. La figure 3.3 explique clairement cette géométrie.

La figure 3.4 illustre deux situations : une où le modèle réduit n'est pas rejeté et une autre où il l'est.

3.10 Prévion

Comme nous l'avons fait au chapitre 2, nous allons ici distinguer l'inférence faite sur la moyenne $E[Y]$ de l'inférence sur l'observation Y .

3.10.1 Inférence concernant $E(Y; \mathbf{x})$

Supposons que l'on veuille estimer la valeur moyenne de la variable endogène pour une combinaison de valeurs $\mathbf{x}^{*\top} = (x_1^*, \dots, x_{p'}^*)$ spécifiées des variables exogènes. On veut donc estimer

$$E(Y; \mathbf{x}^*) = E(\mathbf{x}^{*\top} \boldsymbol{\beta} + \varepsilon) = \mathbf{x}^{*\top} \boldsymbol{\beta}.$$

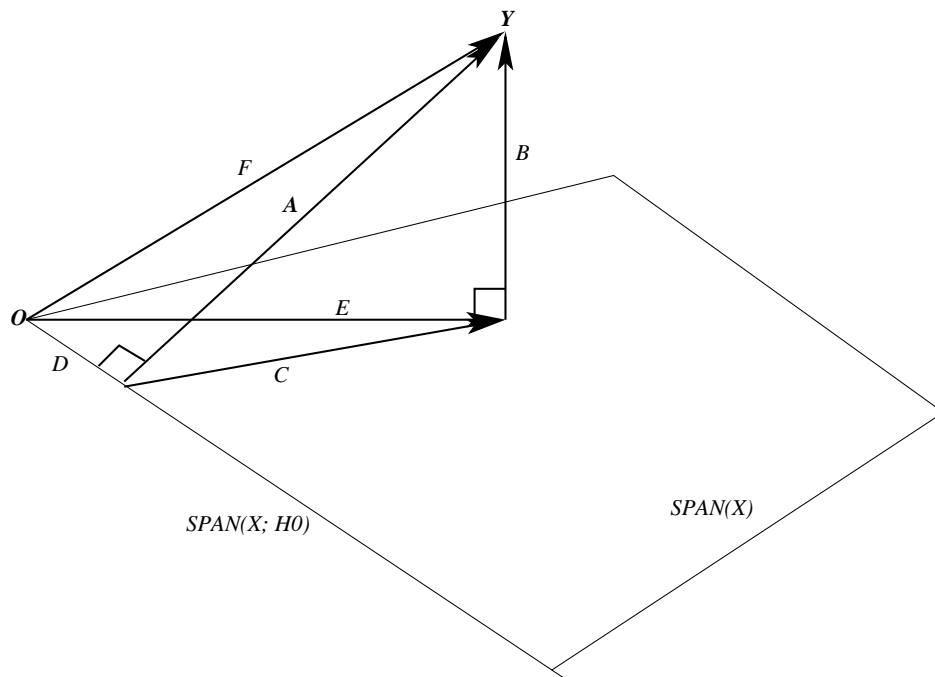


Figure 3.3: Géométrie du principe de somme des carrés résiduels additionnelle. Le plan $SPAN(\mathbf{X})$ représente l'espace colonne de la matrice de schéma sous le modèle complet, tandis que la droite $SPAN(\mathbf{X}; H_0)$ représente l'espace colonne de la matrice d'incidence sous le modèle réduit. La statistique F du test est construite en standardisant les longueurs des vecteurs de la figure en les divisant par la dimension de l'espace dans lesquels ils vivent (leurs degrés de liberté). Le vecteur de longueur E est $\mathbf{X}\hat{\boldsymbol{\beta}}$ sous le modèle complet et vit dans un espace de dimension p . Ce vecteur est la projection orthogonale du vecteur \mathbf{Y} de longueur F sur $SPAN(\mathbf{X})$. Le vecteur de longueur B est le vecteur \mathbf{R} sous le modèle complet et vit dans un espace de dimension $n - p$. Le vecteur de longueur D est le vecteur $\mathbf{X}\hat{\boldsymbol{\beta}}$ sous le modèle réduit et vit dans un espace de dimension $p - \Delta_{dl}$. Ce vecteur est la projection orthogonale du vecteur \mathbf{Y} sur $SPAN(\mathbf{X}; H_0)$. Le vecteur de longueur A est le vecteur \mathbf{R} sous le modèle réduit et vit dans un espace de dimension $n - p + \Delta_{dl}$. Le vecteur de longueur C est la différence entre les vecteurs \mathbf{R} des deux modèles, et donc C^2 est la différence entre les sommes de carrés résiduels. Ce vecteur vit dans un espace de dimension Δ_{dl} . La statistique du test F du principe de somme de carrés résiduels additionnelle est tout simplement égale à $(C^2/\Delta_{dl})/\{B^2/(n - p)\}$.

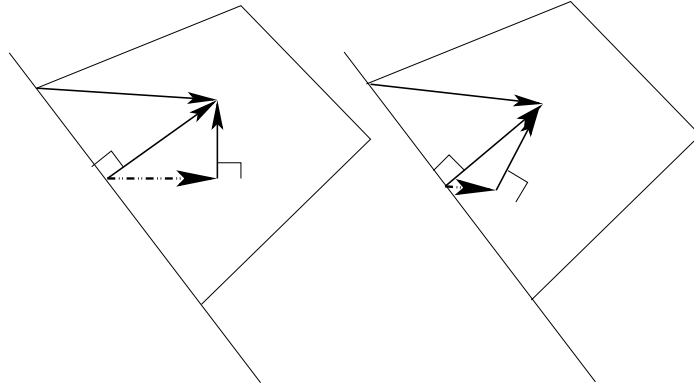


Figure 3.4: Dans la figure de gauche, le modèle réduit serait probablement rejeté, étant donnée la longueur du vecteur pointillé représentant la différence entre les sommes de carrés résiduels. En contrepartie, dans la figure de droite on ne rejeterait probablement pas le modèle réduit.

Le théorème de Gauss–Markov suggère d'utiliser $\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}$. On a

$$\text{var}(\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}) = \mathbf{x}^{*\top} \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{x}^* = \sigma^2 \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*.$$

Puisque $\hat{\boldsymbol{\beta}}$ suit une loi normale, $\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}$ suit aussi une loi normale, et on en déduit donc:

$$\frac{\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} - \mathbf{x}^{*\top} \boldsymbol{\beta}}{\sqrt{s^2 \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*}} \sim t_{n-p}. \quad (3.20)$$

On obtient donc l'intervalle de confiance à $100 \times (1 - \alpha)\%$ pour $E(Y; \mathbf{x}^*)$ donné par

$$\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} \pm t_{\alpha/2; n-p} \sqrt{s^2 \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*}. \quad (3.21)$$

3.10.2 Inférence concernant une valeur de Y étant donné \mathbf{x}

Nous cherchons maintenant à prédire la valeur même d'une réalisation de la variable endogène pour une combinaison de valeurs \mathbf{x}^* fixées des variables exogènes.

En fait on cherche à estimer $Y = \mathbf{x}^{*\top} \boldsymbol{\beta} + \varepsilon$. Le théorème de Gauss–Markov recommande d’estimer $\mathbf{x}^{*\top} \boldsymbol{\beta}$ avec $\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}$. Comme $E(\varepsilon) = 0$, la meilleure estimation d’une réalisation de ε est 0. L’estimation ponctuelle de Y à une valeur \mathbf{x}^* donnée sera donc $\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} + 0 = \mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}$. Afin de pouvoir utiliser un résultat similaire à (3.7), nous devons calculer

$$\begin{aligned} \text{var}\{\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} - (\mathbf{x}^{*\top} \boldsymbol{\beta} + \varepsilon)\} &= \text{var}(\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}}) + \text{var}(\varepsilon) \\ &= \sigma^2 \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^* + \sigma^2 = \sigma^2 \{1 + \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*\}. \end{aligned}$$

On trouve

$$\boxed{\frac{\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} - (\mathbf{x}^{*\top} \boldsymbol{\beta} + \varepsilon)}{\sqrt{s^2 \{1 + \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*\}}} \sim t_{n-p}.} \quad (3.22)$$

Ceci nous mène à l’intervalle de confiance à $100 \times (1 - \alpha)\%$ pour Y étant donnée la combinaison des variables exogènes \mathbf{x}^* , donné par

$$\boxed{\mathbf{x}^{*\top} \hat{\boldsymbol{\beta}} \pm t_{\alpha/2; n-p} \sqrt{s^2 \{1 + \mathbf{x}^{*\top} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}^*\}}.} \quad (3.23)$$

Exemple 3.8 (Consommation d’essence) Retournons au modèle $Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_4 x_{i4} + \varepsilon_i$. Une prévision pour la consommation d’essence par individu en moyenne au cours des années dans un état où la taxe de vente est 8%, le taux de permis de conduire est de 60%, le revenu moyen est 4 000 \$ et où on compte 5 000 miles de routes fédérales est de

$$(1, 8, 60, 4, 5)(377.291146, -34.790149, 13.364494, -66.588752, -2.425889)^\top = 622.$$

L’intervalle de confiance à 95% pour cette prévision est donc

$$\begin{aligned} &622 \pm 2.02 \times \sqrt{4396.511(1, 8, 60, 4, 5)(\mathbf{X}^\top \mathbf{X})^{-1}(1, 8, 60, 4, 5)^\top} \\ &= 622 \pm 2.02 \times \sqrt{4396.511 \times 0.0407} \\ &= (595, 649). \end{aligned}$$

Si on avait demandé une prévision pour la consommation moyenne par individu dans le même état pour une année donnée, alors on aurait la même prévision ponctuelle, mais l'intervalle de confiance serait plutôt

$$622 \pm 2.02\sqrt{4396.511(1 + 0.0407)} = (485, 759).$$