

MAT7381 Solution #5.1

a)

	facture	revenu	personnes	surface
facture	1.000	0.837	0.494	0.905
revenu	0.837	1.000	0.143	0.961
personnes	0.494	0.143	1.000	0.366
surface	0.905	0.961	0.366	1.000

Comme on pouvait s'y attendre, c'est la grandeur de la maison qui influe le plus sur la consommation. Le revenu aussi et fortement corrélé avec la consommation. S'agit-il du revenu comme tel (les riches consomment plus librement, disposent de plus d'appareils, etc.) ? Ou est-ce simplement de fait que les riches ont de plus grande maisons ?

b)

```
> reglin123<-lm(facture~revenu+personnes+surface)
> summary(reglin123)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -358.44157   198.73583  -1.804  0.0813 .
revenu       0.07514     0.13609   0.552  0.5850
personnes    55.08763    29.04515   1.897  0.0675 .
surface      0.28110     0.22611   1.243  0.2234

Residual standard error: 135.4 on 30 degrees of freedom
Multiple R-squared: 0.8514,    Adjusted R-squared: 0.8365
F-statistic: 57.28 on 3 and 30 DF,  p-value: 1.585e-12
```

Le nombre de personnes influence, bien sûr, la facture d'électricité. Mais sa corrélation avec la facture n'est pas si forte car d'autres facteurs jouent un rôle plus important : un ménage de six personnes consomme plus qu'un ménage de deux, mais un ménage de 6 personnes peut vivre dans une grande maison comme dans une petite. Ce qui entraîne une grande variation dans la facture et donc une faible corrélation. Dans la régression multiple, le coefficient de x_2 représente le taux d'accroissement de y par rapport à x_2 lorsque les autres variables exogènes restent fixes. Ce qu'on conclut ici, c'est que pour des maisons de même taille, le nombre de personnes influe sur la consommation d'électricité.

c)

```
> reglin23<-lm(facture~personnes+surface)
> summary(reglin23)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -255.94923    70.14230  -3.649 0.000959 ***
personnes    42.03394    16.67947   2.520 0.017096 *
surface      0.40429     0.03615  11.183 2.08e-12 ***

Residual standard error: 133.9 on 31 degrees of freedom
Multiple R-squared: 0.8499,    Adjusted R-squared: 0.8402
F-statistic: 87.74 on 2 and 31 DF,  p-value: 1.72e-13
```

Les deux variables se révèlent du coup significatives. Le fait que la surface n'était pas significative en présence de « revenu » est dû à la forte corrélation entre la surface et le revenu : quand le revenu est connu, la surface n'ajoute pas grand-chose comme information (celle-ci est contenue dans « revenu »).

d)

```
> reglin1<-lm(facture~revenu)
> summary(reglin1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -425.16251  124.92953  -3.403  0.00181 **
revenu       0.25899    0.02995   8.649 6.97e-10 ***

Residual standard error: 186.2 on 32 degrees of freedom
Multiple R-squared: 0.7004,    Adjusted R-squared: 0.691
F-statistic: 74.8 on 1 and 32 DF,  p-value: 6.974e-10
```

Deux hypothèses possibles : 1) le revenu ne détermine pas la consommation, mais il est indirectement lié à la consommation, de par sa relation avec la taille de la maison; 2) Le revenu *en soi* a un effet sur la consommation; en d'autres termes, dans des maisons de même taille, celles avec un revenu élevé consommeront davantage. La 2^e hypothèse n'est pas appuyée par les données : lorsque la surface est l'une des variables exogènes, le revenu est non significatif. Ce qui ne veut pas dire, cependant, qu'elle est fausse. On n'a pas démontré que le coefficient de « revenu » est nul : nous affirmons seulement qu'il *pourrait* l'être. À ce stade, la conclusion suivante peut être acceptée provisoirement : le revenu a possiblement un effet sur la consommation, mais cet effet est mineur comparé à celui qu'il exerce par l'entremise de la taille de la maison.