MAT7381 Solution #5.2

a)

```
> reglin123<-lm(y~age+poids +pli)
> summary(reglin123)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.707875   0.326839   5.225 7.02e-06 ***
age         -0.002150   0.003268  -0.658  0.51464
poids       -0.015210   0.004771  -3.188  0.00291 **
pli          0.202381   0.166465   1.216  0.23178

Residual standard error: 0.2152 on 37 degrees of freedom
Multiple R-squared: 0.3387,     Adjusted R-squared: 0.2851
F-statistic: 6.317 on 3 and 37 DF,  p-value: 0.001436
```

La variable « âge » ne semble nettement pas pertinente en présence des deux autres.

b)

```
> reglin3<-lm(y~pli)
> summary(reglin3)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.7132     0.1220   5.848 8.41e-07 ***
pli          -0.1502     0.1576  -0.953    0.347

Residual standard error: 0.2549 on 39 degrees of freedom
Multiple R-squared: 0.02275,    Adjusted R-squared: -0.002312
F-statistic: 0.9077 on 1 and 39 DF,  p-value: 0.3466
```

La variable “pli” n’est pas significative. On pourrait donc songer à l’éliminer de la régression.

c)

```
> reglin1<-lm(y~age)
> summary(reglin1)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.497088   0.350574   4.270 0.000121 ***
age         -0.007859   0.003066  -2.564 0.014325 *

Residual standard error: 0.2385 on 39 degrees of freedom
Multiple R-squared: 0.1442,     Adjusted R-squared: 0.1223
F-statistic: 6.572 on 1 and 39 DF,  p-value: 0.01433
```

Contrairement à la conclusion tirée en a), la variable « âge » peut contribuer à prédire *y* en l’absence des autres variables. Non seulement cela, elle semble préférable à la variable « pli » alors que dans la première analyse elle s’était révélée moins utile.

d)

```
> summary(lm(y~age+poids))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.737292   0.327986   5.297 5.24e-06 ***
age         -0.003118   0.003189  -0.978  0.33448
poids       -0.011761   0.003859  -3.048  0.00418 **

Residual standard error: 0.2166 on 38 degrees of freedom
Multiple R-squared: 0.3123,     Adjusted R-squared: 0.2761
F-statistic: 8.628 on 2 and 38 DF,  p-value: 0.0008139
```

L’« âge », d’un coup, semble perdre de son utilité à la faveur du poids.

e) Le tableau des corrélations peut aider ici :

```
> round(cor(Acide),3)
       acide    age  poids    pli
acide  1.000 -0.380 -0.543 -0.151
age   -0.380  1.000  0.488  0.101
poids -0.543  0.488  1.000  0.566
pli   -0.151  0.101  0.566  1.000
```

La variable la plus fortement corrélée avec *y* est le poids (*r* = -0,543). Donc à elle seule, elle prédit *y* mieux que l'âge (*r* = -0,380). La (relativement) forte corrélation entre l'âge et le poids est ce qui cause la perte d'influence de la première en présence de la dernière. Une hypothèse est qu'en fait c'est le poids seul qui joue, et que si l'âge est corrélée avec *y*, c'est parce qu'elle est corrélée avec le poids : les adolescents plus âgés pèsent plus et font décroître *y*. En termes plus formels, le coefficient de « âge » dans une régression qui compte l'âge *et* le poids comme variables exogènes, est le taux d'accroissement (décroissement, ici) de *y* par rapport à l'âge *lorsque le poids reste fixe*. On conclut (provisoirement) que pour un poids donné, l'âge

Est-ce qu'on devrait laisser tomber l'âge ? Le coefficient de détermination (R^2) augmente de 0,295 (sans l'âge) à 0,312 avec l'âge. La différence est négligeable, on retient donc pour le moment l'hypothèse que l'âge en soi ne joue pas et on se contente de ne retenir que le poids comme variable exogène.