Q.N. 4) Table below presents the results of a study in which the lung capacities of 16 children were measured. Lung capacity is measured by having a person take as deep a breath as possible, then blowing into a tube connected to a machine called a spirometer, which measures the volume of air exhaled. Variables that might be useful to predict a child's lung capacity are height, weight, and age, along with the temperature and barometric pressure at the time the lung capacity was measured. In the table lung capacity (y) is measured in liters, height (x_1) in inches, weight (x_2) in pounds, age (x_3) in years, barometric pressure (x_4) in inches, and temperature (x_5) in degrees Fahrenheit. (The data is also provided in the Blackboard Exam 2–4)

y	x1	x2	x3	x4	x5
2.68	61	105	13	30.0	70.4
2.22	49	50	9	30.3	71.6
2.29	61	128	15	29.9	73.6
2.16	54	81	10	29.6	70.8
3.19	68	120	13	29.6	69.1
1.97	56	92	10	30.0	70.1
2.98	68	140	16	30.2	67.7
1.96	57	86	10	30.7	56.3
3.05	62	120	16	30.5	63.8
2.48	60	114	13	30.3	69.9
2.09	57	94	12	30.1	73.0
3.27	67	145	16	30.6	70.3
2.26	52	66	10	28.8	75.5
3.20	72	159	16	30.0	72.1
2.10	53	65	10	30.5	67.5
1.79	51	56	9	29.7	65.6

a) Fit a multiple linear regression model and determine the coefficient of determination.
b) Construct 90% confidence interval for model parameters.
c) Identify all the significant variables by performing the marginal t-test.
d) Check for multicollinearity among the regressor variables.
e) Predict the lung capacity for a 12-year-old who is 61 inches tall and weighs 105 pounds, at a pressure of 30.1 inches and a temperature of 70 degrees.

Solution: We saved the data and imported in R and the data is displayed as below

```
> data
      y x1  x2 x3   x4   x5
1  2.68 61 105 13 30.0 70.4
2  2.22 49  50  9 30.3 71.6
```

```
3   2.29 61 128 15 29.9 73.6
4   2.16 54  81 10 29.6 70.8
5   3.19 68 120 13 29.6 69.1
6   1.97 56  92 10 30.0 70.1
7   2.98 68 140 16 30.2 67.7
8   1.96 57  86 10 30.7 56.3
9   3.05 62 120 16 30.5 63.8
10 2.48 60 114 13 30.3 69.9
11 2.09 57  94 12 30.1 73.0
12 3.27 67 145 16 30.6 70.3
13 2.26 52  66 10 28.8 75.5
14 3.20 72 159 16 30.0 72.1
15 2.10 53  65 10 30.5 67.5
16 1.79 51  56  9 29.7 65.6
```

a) We used R code below to fit a multiple linear regression model

```
> model=lm(y~x1+x2+x3+x4+x5)
> model

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5)

Coefficients:
(Intercept)       x1       x2       x3       x4       x5
   -7.81721   0.11927  -0.02463  0.16411  0.07640  0.02022
```

Therefore the fitted model is:

Lung capacity = -7.81720523 + 0.11927217*height - 0.02462661*weight + 0.16411312*age + 0.07639732*barometric_pressure + 0.02021925*temperature

```
> summary(model)$r.sq

[1] 0.87955
```

The coefficient of determination is 0.87955

b) We used R code below to construct 90% confidence interval for the model parameters

```
> confint(model, level=0.9)
                  5 %       95 %
(Intercept) -19.21319849 3.578788022
x1            0.05961491 0.178929419
x2           -0.04087617 -0.008377045
```

x3 0.05079086 0.277435371

x4 -0.22169633 0.374490963

x5 -0.01088270 0.051321200

Therefore, 90% confidence intervals for β_0, β_1, β_2, β_3, β_4 and β_5 are (-19.213, 3.579), (0.0596, 0.1789), (-0.0409, -0.0084), (0.0508, 0.2774), (-0.2217, 0.3745) and (-0.0109, 0.0513) respectively.

c) We will perform the marginal t-test using R- code below

```
> summary(model)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32133 -0.14651 -0.00255  0.14476  0.28185

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.917205   6.287579  -1.243  0.24212
x1           0.119272   0.032915   3.624  0.00466 **
x2          -0.024627   0.008965  -2.747  0.02059 *
x3           0.164113   0.062524   2.625  0.02539 *
x4           0.076397   0.164469   0.465  0.65223
x5           0.020219   0.017160   1.179  0.26597
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observe that the p-value for x1, x2 and x3 is less than 0.05. This means height, weight, and age are significant variables at 0.05 level of significance.

d) We used R code below to check for multicollinearity

```
> library(car)
> vif(model)
      x1       x2       x3       x4       x5
16.556535 28.582097  9.496716  1.999309  1.970648
```

Note that three of the variance inflation factors 16.556535, 28.582097 and 9.496716 are fairly large indicating that the variance of the estimated coefficients of x1, x2 and x3 is inflated means these variables are correlated with at least one of the other predictors in the model.

e) *We use R code to make the prediction*

```
> predict(model, data.frame(x1 = 61, x2 = 105, x3 = 12, x4
= 30.1, x5 = 70))
       1
2.556867
```

Hence, the lung capacity for a 12-year-old who is 61 inches tall and weighs 105 pounds, at a pressure of 30.1 inches and a temperature of 70 degrees is `2.556867 liters.`