Part A

```
> lm_fit=lm(newSiriBFperc~Weight+Height+NeckC+ChestC+AbdomenC+HipC+ThighC+Kne
eC+AnkleC+BicepsC+ForearmC+WristC+dummy)
> summary(lm_fit)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)

dummy        0.10052    0.69435   0.145  0.88502

> lm_1=update(lm_full,.~.-Over45)
> summary(lm_1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
KneeC        0.07088    0.24386   0.291  0.77157
---

> lm_2=update(lm_1,.~.-KneeC)
> summary(lm_2)
Coefficients:
ThighC       0.08890    0.13818   0.643  0.52063
---

> lm_3=update(lm_2,.~.-ThighC)
> summary(lm_3)
Coefficients:
Weight      -0.03504    0.06383  -0.549 0.583485
---

> lm_4=update(lm_3,.~.-Weight)
> summary(lm_4)
Coefficients:
ForearmC     0.20957    0.20698   1.013 0.312311
---

> lm_5=update(lm_4,.~.-ForearmC)
> summary(lm_5)
Coefficients:
AnkleC       0.23520    0.21275   1.106  0.27003
---

> lm_6=update(lm_5,.~.-AnkleC)
> summary(lm_6)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.96118    8.23097   2.547  0.01150 *
Height      -0.37610    0.12712  -2.959  0.00340 **
NeckC       -0.36516    0.22459  -1.626  0.10527
ChestC      -0.18595    0.09065  -2.051  0.04130 *
AbdomenC     0.99730    0.07767  12.840  < 2e-16 ***
HipC        -0.19133    0.10123  -1.890  0.05995 .
BicepsC      0.33298    0.15265   2.181  0.03012 *
WristC      -1.46315    0.46914  -3.119  0.00204 **
---
```

Answer: From The each step to regression, we can compare the largest Pvalue to delete from the model, the order is dummy(Over45),Kneec, ThighC, Weight ,ForearmC,AnkleC

The regression model is newSiriBFperc=20.96118-0.37610Height-0.36516NeckC-0.18595ChestC+0.99730 AbdomenC-0.19133 HipC+0.33298 BicepsC-1.46315 WristC

Part B

```
> lm_empty=lm(Bodyfate$newSiriBFperc~1)
> summary(lm_empty)

Residuals:
     Min       1Q   Median       3Q      Max
-18.9859  -6.5420   0.2306   6.2217  28.5015

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.9859     0.5286   35.92   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.357 on 249 degrees of freedom


> pvalue_for1=numeric(13)
> names(pvalue_for1)=c("Weight","Height","NeckC","ChestC","AbdomenC","HipC","
ThighC","KneeC","AnkleC","BicepsC","ForearmC","WristC","Over45")
Weight       Height        NeckC       ChestC     AbdomenC         HipC
1.105322e-25 4.464991e-01 3.452486e-15 1.834532e-35 1.027911e-59 1.198462e-27
      ThighC        KneeC       AnkleC      BicepsC     ForearmC       WristC
2.473884e-20 1.536127e-15 1.106314e-04 1.910743e-15 9.617845e-09 3.388341e-07
      Over45
3.279761e-02


> lm_for1=update(lm_empty,.~.+AbdomenC)
> summary(lm_for1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.3067     2.8230  -14.99   <2e-16 ***
AbdomenC      0.6642     0.0304   21.84   <2e-16 ***
```

```
> pvalue_for2=numeric(12)
> names(pvalue_for2)=c("Weight","Height","NeckC","ChestC","HipC","ThighC","Kn
eeC","AnkleC","BicepsC","ForearmC","WristC","Over45")
      Weight       Height        NeckC       ChestC         HipC       ThighC
2.354923e-10 1.730149e-08 1.407977e-06 1.857095e-04 1.439466e-05 2.112094e-02
       KneeC       AnkleC      BicepsC     ForearmC       WristC       Over45
1.022381e-04 8.833134e-03 3.876392e-02 1.398205e-02 1.697697e-09 4.905053e-01
> lm_for2=update(lm_for1,.~.+Weight)
> summary(lm_for2)

Call:
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -47.41951    2.71976  -17.43  < 2e-16 ***
AbdomenC      0.99759    0.05774   17.28  < 2e-16 ***
Weight       -0.14408    0.02180   -6.61 2.35e-10 ***
---


> pvalue_for3=numeric(11)
> names(pvalue_for3)=c("Height","NeckC","ChestC","HipC","ThighC","KneeC","Ank
leC","BicepsC","ForearmC","WristC","Over45")
      Height        NeckC       ChestC         HipC       ThighC        KneeC
0.1094944470 0.0495148635 0.3343920729 0.7687362450 0.0282540883 0.8630623544
      AnkleC      BicepsC     ForearmC       WristC       Over45
0.5455442304 0.0529244274 0.3015792059 0.0005940908 0.0827633997
> lm_for3=update(lm_for2,.~.+WristC)
> summary(lm_for3)


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -25.23010    6.90997  -3.651 0.000319 ***
AbdomenC      0.97919    0.05673  17.261  < 2e-16 ***
Weight       -0.09985    0.02483  -4.022 7.69e-05 ***
WristC       -1.55701    0.44748  -3.479 0.000594 ***
---



> pvalue_for4=numeric(10)
> names(pvalue_for4)=c("Height","NeckC","ChestC","HipC","ThighC","KneeC","Ank
leC","BicepsC","ForearmC","Over45")
    Height      NeckC     ChestC       HipC     ThighC      KneeC     AnkleC
0.08753103 0.43136745 0.44007787 0.73179176 0.12301391 0.56900979 0.17407271
   BicepsC   ForearmC     Over45
0.02001777 0.07908759 0.59822035
> lm_for4=update(lm_for3,.~.+BicepsC)
> summary(lm_for4)

Call:
lm(formula = Bodyfate$newSiriBFperc ~ AbdomenC + Weight + WristC +
    BicepsC)
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -30.45503    7.20232  -4.229 3.32e-05 ***
AbdomenC      0.99108    0.05645  17.558  < 2e-16 ***
Weight       -0.13214    0.02821  -4.685 4.65e-06 ***
WristC       -1.65671    0.44550  -3.719 0.000248 ***
BicepsC       0.36298    0.15503   2.341 0.020018 *
---
> pvalue_for5=numeric(9)
> names(pvalue_for5)=c("Height","NeckC","ChestC","HipC","ThighC","KneeC","Ank
leC","ForearmC","Over45")
     Height       NeckC      ChestC        HipC      ThighC       KneeC
 0.26942476 -0.28506613  0.30902147  0.73958012  0.32316175  0.49763072
     AnkleC    ForearmC      Over45
```

0.12863924 0.30331958 0.02168526

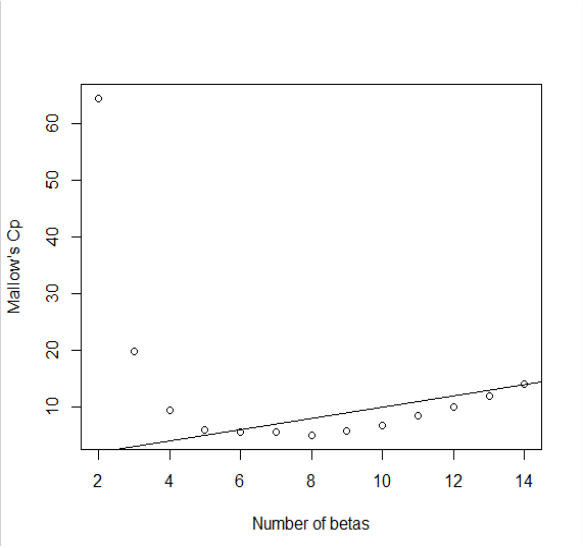
We first write the empty model, and computer each Pvaule of variable, we choose the least Pvalue variable to the empty model in each step, the order is AbdomenC ,Weight ,WristC,BicepsC
The regression model is newSiriBFperc =0.99108AbdomenC -0.13214Weight -1.6567 1 WristC+0.36298BicepsC-30.45503

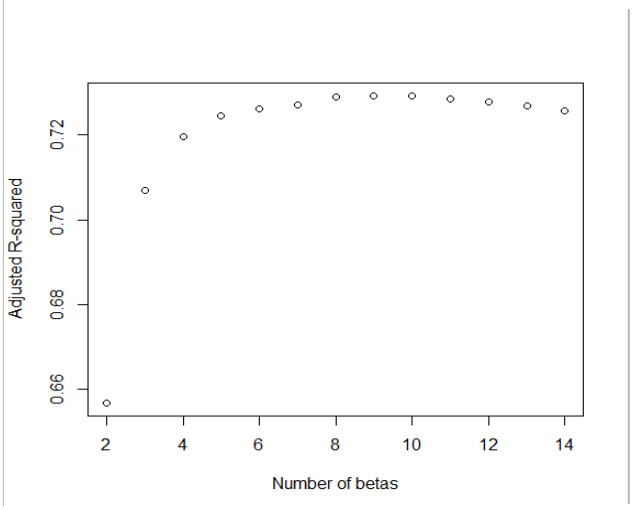
Partc

```
> subset=regsubsets(newSiriBFperc~Weight+Height+NeckC+ChestC+AbdomenC+HipC+Th
ighC+KneeC+AnkleC+BicepsC+ForearmC+WristC+Over45,method="exhaustive",nbest=1,
nvmax=13,data=Bodyfate)
> sum_subset$which
```

	(Intercept)	Weight	Height	NeckC	ChestC	AbdomenC	HipC	ThighC	KneeC	AnkleC	BicepsC	ForearmC	WristC	Over45
1	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
4	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
5	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
6	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
7	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
8	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
9	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
10	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
11	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE
12	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE
13	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

```
> R2_adj
 [1] 0.6566308 0.7070561 0.7196620 0.7246780 0.7261585 0.7272627 0.7290949
 [8] 0.7293434 0.7293717 0.7285817 0.7279145 0.7268638 0.7257308
> C_p
 [1] 64.481671 19.818005  9.443323  5.940477  5.619517  5.642793  5.031670
 [8]  5.825575  6.813997  8.515676 10.105091 12.020958 14.000000
```

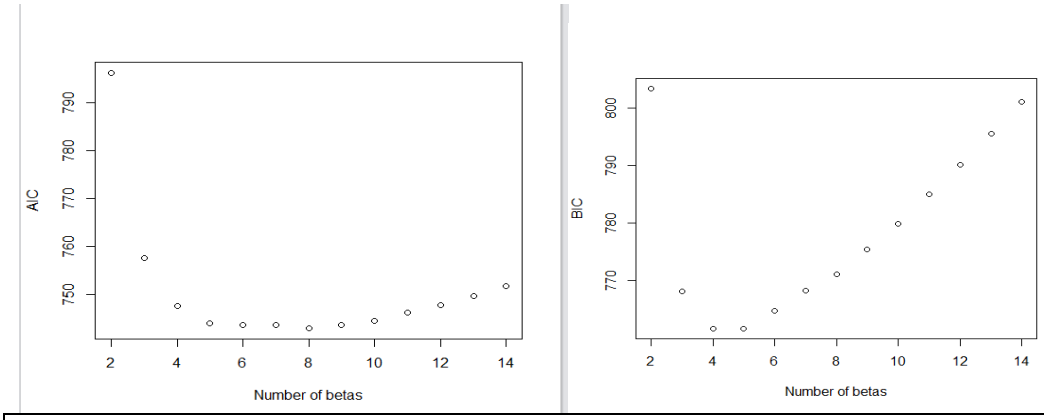





```
> aic_p
```

```
 [1] 796.3263 757.6099 747.5994 744.0674 743.6969 743.6602 742.9440 743.6794
 [9] 744.6138 746.2987 747.8643 749.7752 751.7530

> bic_p
 [1] 803.3692 768.1743 761.6653 761.6747 764.8257 768.3104 771.1157 775.3726
 [9] 779.8284 785.0347 790.1218 795.5542 801.0534
```



For AIC, we choose model 7(the AIC IS LEAST) 7 predictors are: `Height NeckC ChestC AbdomenC HipC ForearmC WristC`.

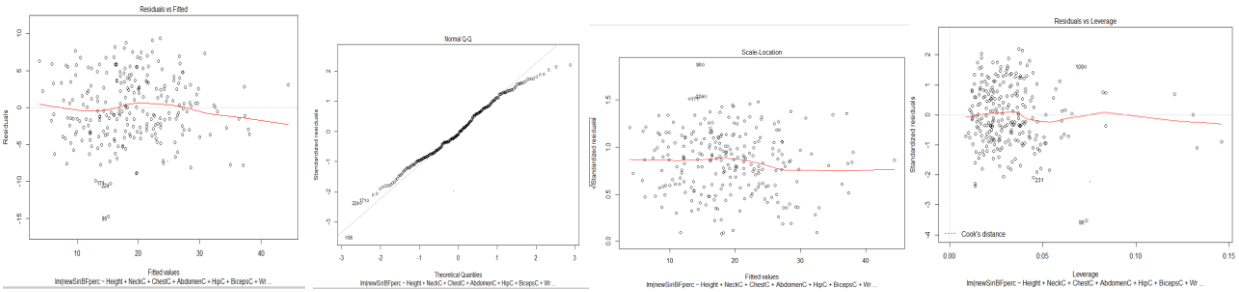
For BIC, we choose model 3(the BIC is the least value). There are 4 predictors: `Weight AbdomenC WristC`

For CP,we choose model 5, because the variable is least and approach 5.There are 5 predictors: `Weight AbdomenC AnkleC BicepsC WristC`

For adjust R^2, we choose model 9, the value of R^2 adjust is the highest. There are 9 predictors: `Height NeckC ChestC AbdomenC HipC ThighC AnkleC BicepsC ForearmC WristC`

`For Aic choosen model, the R^2 adjusted is very close to` adjust R^2 choosen model, but the predictor is small than that ,it much better.
For the cp of BIC model also close to P=4, (it is less close compared to CP best model).But the predictors is less, might be suitable for simplistics.

Part D
For Amodel

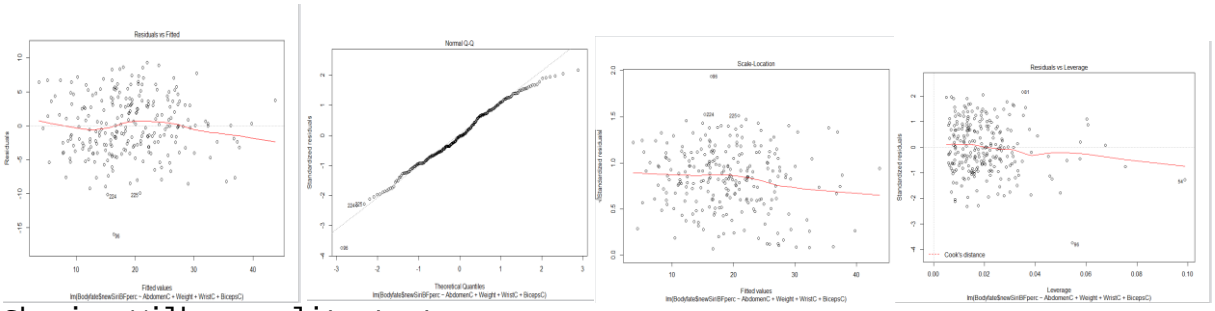


```
Shapiro-Wilk normality test

data:  lm.backward$residuals
```

```
W = 0.99289, p-value = 0.2775
```

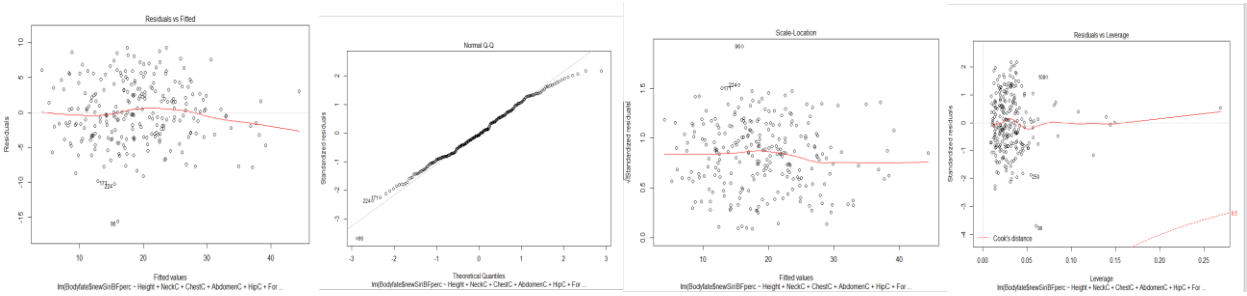
For model B




```
Shapiro-Wilk normality test

data:  lm.forward$residuals
W = 0.98869, p-value = 0.04722
```

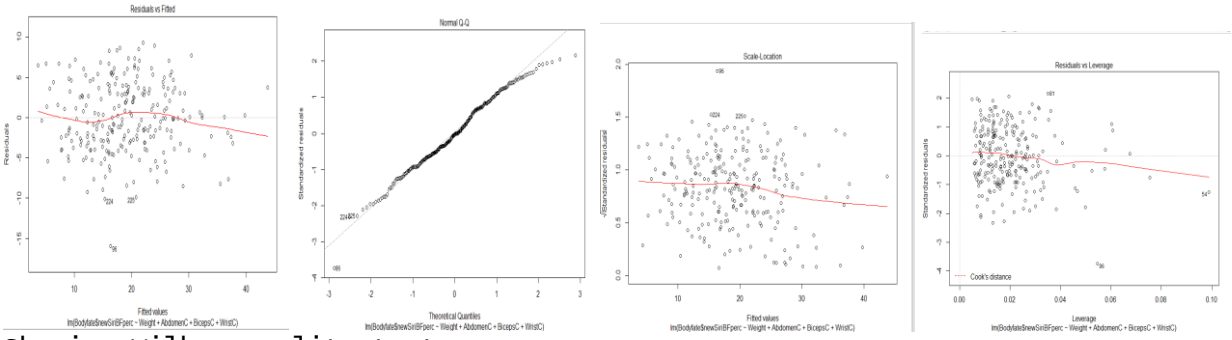
AiC model




```
Shapiro-Wilk normality test

data:  lm.aic.p$residuals
W = 0.99289, p-value = 0.2775
```

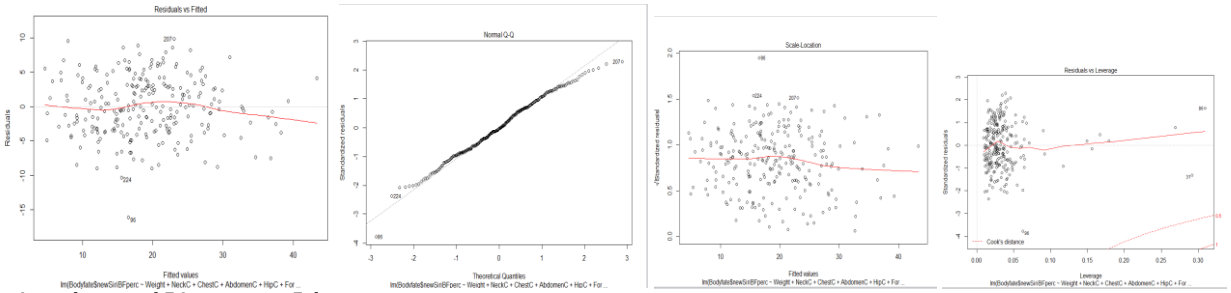
BIC model




```
Shapiro-Wilk normality test

data:  lm.aic.p$residuals
W = 0.98886, p-value = 0.05095
```

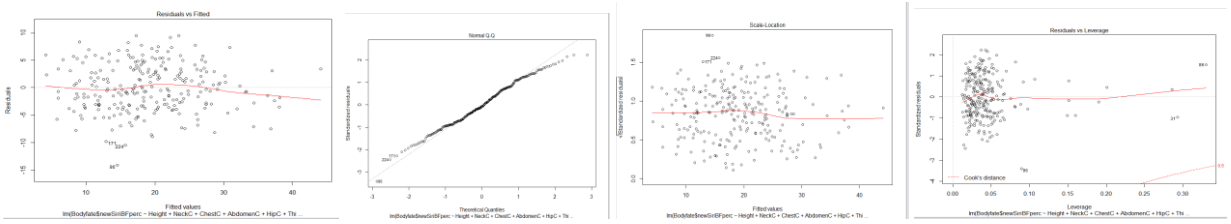
Cp




```
Shapiro-Wilk normality test

data:  lm.bic.p$residuals
W = 0.9889, p-value = 0.0517
```

Adjustd R^2




```
Shapiro-Wilk normality test

data:  lm.R2.adj$residuals
W = 0.99292, p-value = 0.2809
```

A model Vif

```
Height    NeckC   ChestC AbdomenC     HipC  BicepsC   WristC
1.456085 3.482255 7.188742 8.272218 5.599539 2.618959 2.419499
```

B model Vif:

```
AbdomenC   Weight   WristC  BicepsC
4.298628 7.529088 2.146810 2.657965
```

ACI model VIF

```
Height    NeckC   ChestC AbdomenC     HipC ForearmC   WristC
1.456617 3.509739 7.112527 8.379800 5.333024 2.041229 2.457565
```

```
Bic MODEL VIF
Weight AbdomenC   WristC
5.729221 4.263829 2.127194
```

```
R^2 adjusted
Height    NeckC   ChestC AbdomenC     HipC   ThighC   AnkleC  BicepsC Forearm
C   WristC
1.495249 3.705737 7.463192 8.519283 9.545111 5.474892 1.654904 3.271572 2.330
677 2.738371
```

```
Cp moldel VIF
Weight     NeckC    ChestC  AbdomenC      HipC  ForearmC    AnkleC    WristC
18.225219  3.793135  8.186659  7.965648  9.355490  2.129133  1.723900  2.6627
70
```

From this VIF, we see the model BIC's vif is less than others(the biggest one in the BIC model), and the predictor is less than others. Also from the plot, we can also see the normal distribution and others are good, I will choose the predictor `Weight,AbdomenC,WristC`

```
> lmfinal=lm(Bodyfate$newSiriBFperc ~ Weight+AbdomenC+WristC,data=Bodyfate)
> summary(lmfinal)

Call:
lm(formula = Bodyfate$newSiriBFperc ~ Weight + AbdomenC + WristC,
    data = Bodyfate)


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -25.23010    6.90997  -3.651 0.000319 ***
Weight       -0.09985    0.02483  -4.022 7.69e-05 ***
AbdomenC      0.97919    0.05673  17.261  < 2e-16 ***
WristC       -1.55701    0.44748  -3.479 0.000594 ***
---
```

The regression model is Bodypercent=-25.23010+0.97919 AbdomenC-0.09985 Weight -1.55701WristC

The interpretation of the model

Holind other variables constant, when weight increase 1 unit, the bodypercent decrease -0.09985 and this variables sd is 0.02483
Holind other variables constant, when AbdomenC increase 1 unit, the bodypercent increase 0.97919 and this variables sd is 0.05673
Holind other variables constant, when WristC increase 1 unit, the bodypercent decrease -1.55701 and this variables sd is 0.44748
The intercept of the model has no real meaning in the real life, because the bodypercent can not be negative, and weight can not be 0

For 95% CI

```
> predictors=data.frame(weight=mean(bodyfat$Weight),AbdomenC=mean(bodyfat$Abd
omenC),WristC=mean(bodyfat$WristC))
> PI=predict(lm.bic.p,new=predictors,interval='prediction',level=0.95)
> PI
      fit      lwr      upr
 1 18.9852 9.985557 27.98484
```

(First compute the mean for each predictors, and use the prediction to predict the response of man average bodypercent)