**Application**

**A.**

**Hypothesis Test on Weight:**

Call:

lm(formula = bodyfat.percent ~ Weight, data = bodyfat)

Residuals:

Min 1Q Median 3Q Max

-27.1676 -4.6126 0.0375 4.9613 20.9494

Coefficients:

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

(Intercept) -13.94208 2.83363 -4.92 1.58e-06 \*\*\*

Weight 0.18490 0.01573 11.75 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.712 on 248 degrees of freedom

Multiple R-squared: 0.3577, Adjusted R-squared: 0.3551

F-statistic: 138.1 on 1 and 248 DF, p-value: < 2.2e-16

> confint(lm.bodyfat.W, level=0.99)

0.5 % 99.5 %

(Intercept) -21.2976226 -6.5865326

Weight 0.1440604 0.2257362

H0: βweight = 0 H1: βweight ≠ 0

Test statistic is t distribution that t = 11.75 with df=248

p-value for test statistic is smaller than 2 × 10-16.

Since p-value is < α = 0.05, we reject null hypothesis.

Thus, we can conclude that Weight has impact on the body fat percentage.

For every 1 pound increase in Weight, we are 99% confident that there will be an increase of 0.1440604 to 0.2257362 percent in body fat percentage.

**Hypothesis Test on Height:**

Call:

lm(formula = bodyfat.percent ~ Height, data = bodyfat)

Residuals:

Min 1Q Median 3Q Max

-19.3423 -6.5537 0.2821 6.2142 27.5375

Coefficients:

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

(Intercept) 29.8863 14.2522 2.097 0.037 \*

Height -0.1551 0.2026 -0.765 0.445

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.365 on 248 degrees of freedom

Multiple R-squared: 0.002357, Adjusted R-squared: -0.001666

F-statistic: 0.5858 on 1 and 248 DF, p-value: 0.4448

> confint(lm.bodyfat.H, level=0.99)

0.5 % 99.5 %

(Intercept) -7.1095881 66.8822117

Height -0.6809313 0.3708134

H0: βheight = 0 H1: βheight ≠ 0

Test statistic is t distribution that t = -0.765 with df=248

p-value for test statistic is 0.445.

Since p-value is > α = 0.05, we do not reject null hypothesis.

Thus, we can conclude that Height does not have impact on the body fat percentage.

Since Height does not have any impact on body fat percentage, body fat percentage will not change with any increase of Height.

**Hypothesis Test on Abdomen Circumference:**

Call:

lm(formula = bodyfat.percent ~ AbdomenC, data = bodyfat)

Residuals:

Min 1Q Median 3Q Max

-23.1760 -3.5408 0.2143 3.1793 12.8435

Coefficients:

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

(Intercept) -42.29941 2.82409 -14.98 <2e-16 \*\*\*

AbdomenC 0.66407 0.03042 21.83 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.899 on 248 degrees of freedom

Multiple R-squared: 0.6578, Adjusted R-squared: 0.6564

F-statistic: 476.7 on 1 and 248 DF, p-value: < 2.2e-16

> confint(lm.bodyfat.A, level=0.99)

0.5 % 99.5 %

(Intercept) -49.6301857 -34.9686421

AbdomenC 0.5851118 0.7430221

H0: βabodmenc = 0 H1: βabdomenc ≠ 0

Test statistic is t distribution that t = 21.83 with df=248

p-value for test statistic is smaller than 2 × 10-16.

Since p-value is < α = 0.05, we reject null hypothesis.

Thus, we can conclude that Abdomen Circumference has impact on the body fat percentage.

For every 1 cm increase in Abdomen Circumference, we are 99% confident that there will be an increase of 0.5851118 to 0.7430221 percent in body fat percentage.

**Comparing the 3 regressions, we can see that regression of Abdomen Circumference has that largest R2** **and regression of Height has the smallest R2**. **Even though regression of Weight does not fit the data as well as that of Abdomen Circumference, it still fits pretty well to the data. However, the regression of Height has very small R2, which indicates that the regression of Height fits the data poorly. Also, from confident intervals, we can see that both interval of β of Abdomen Circumference and β of Weight do not contain 0 while interval of β of Height contain 0. Thus, Abdomen Circumference and Weight have significant impact on body fat percentage while Height does not have significant impact. This result meets the same conclusion with the hypothesis tests.**

**B.**

> T=(summary.full$coefficients[2,1]-0.5)/summary.full$coefficients[2,2]

> T

[1] 6.351477

> pvalue=pt(T,df=n-p)

> pvalue

[1] 1

Computed from R, we have t=6.351477 with p-value ≈ 1

**Hypothesis Test**:

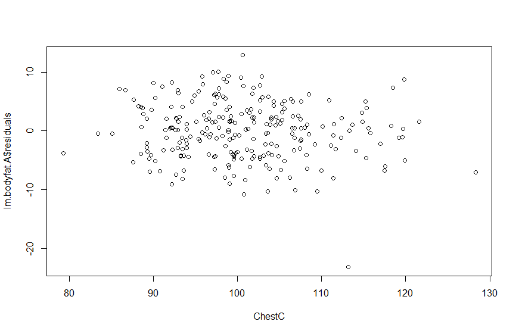
H0: βabodmenc ≥ 0.5 H1: βabdomenc < 0.5

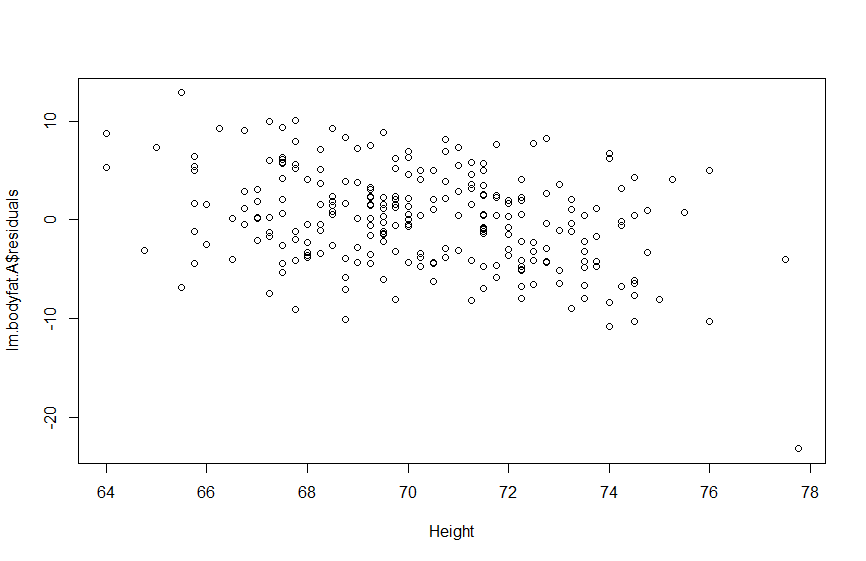
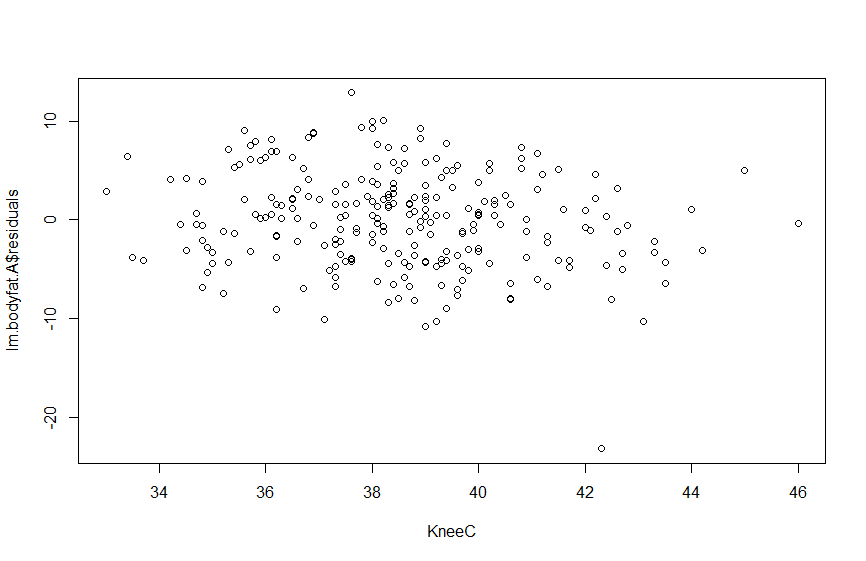
Test statistic is t distribution that t = 6.351477 with df=237

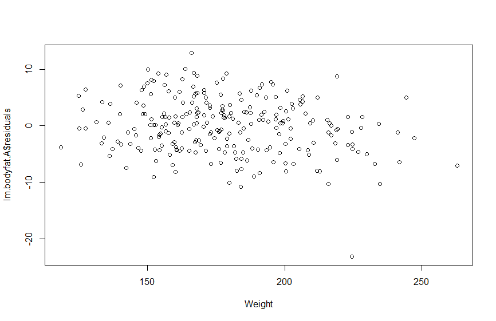
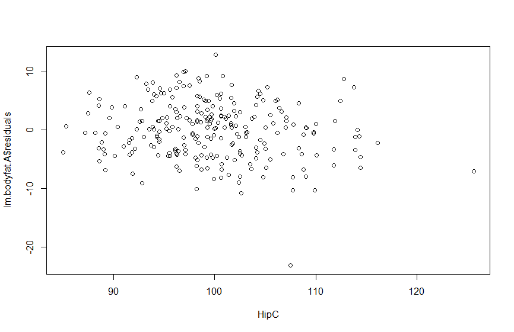
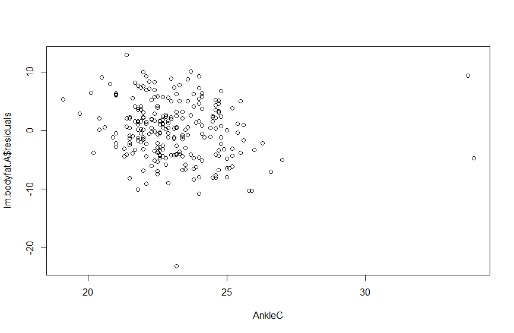
p-value for test statistic is approximately 1.

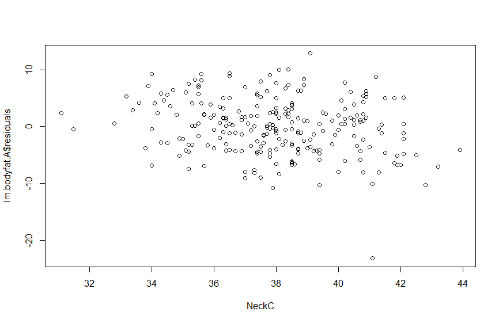
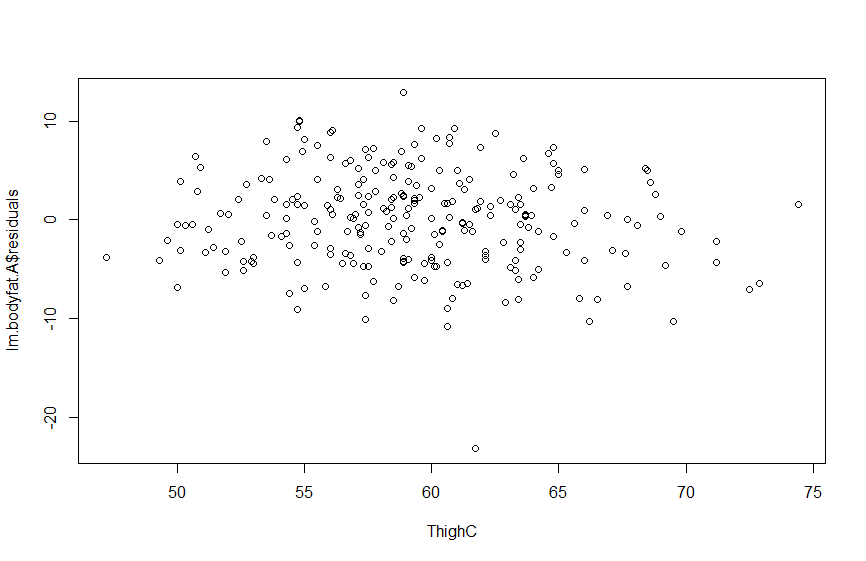
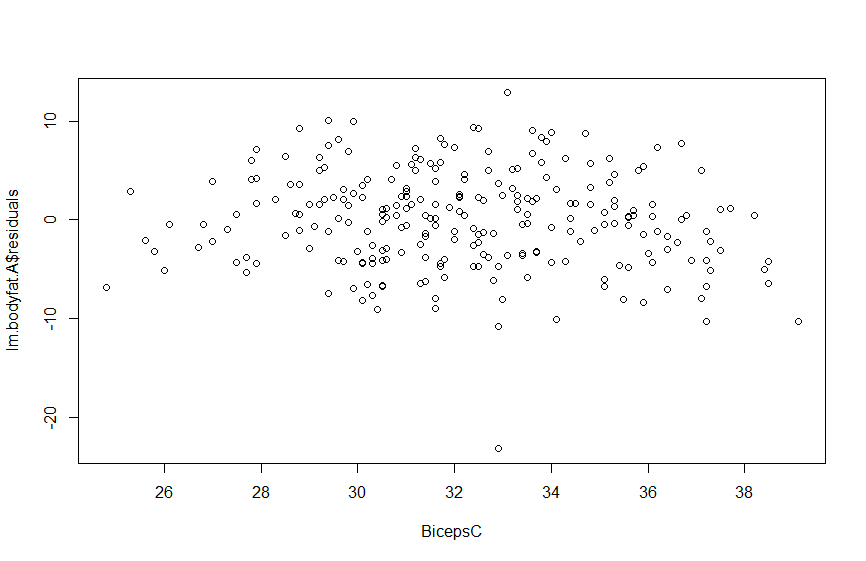
Since p-value is > α = 0.05, we accept null hypothesis.

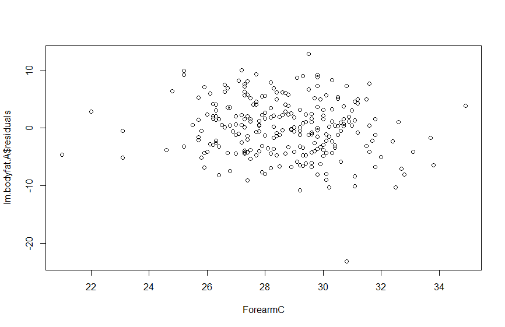
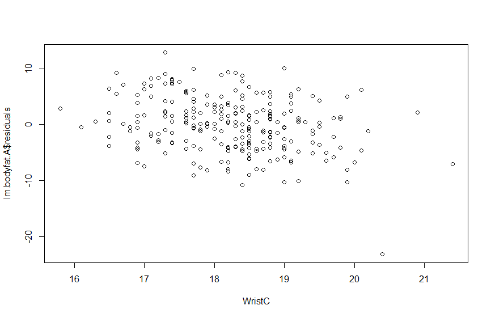
Thus, we can conclude that for each additional cm of abdomen circumference, the body fat percentage increases by more than 0.5 points

 **C.**









By observing the scatter plots, if we find some predictors that have correlation with the residuals of linear regression model for Abdomen Circumference, we can try to put those predictors in our AbdomenC model and test if those added predictors have significant impact on body fat percentage.

From the plots we have above, we can see that Ankle Circumference may have an correlation with residuals since we can see a trend in the plot of AnkleC, which means it may capture some factors that Abdomen Circumference may not where those factors can reduce the residuals of the fitted value. This makes sense because ankle circumference may have influence on the frequency and severity of body movement, which could influence body fat percentage.

**D.**

**Equation of the population model employed**:

Y = β0 + β1X1 + β2X2 + β3X3 + β4X4 + β5X5 + β6X6 + β7X7 + β8X8 + β9X9 + β10X10 + β11X11 + β12X12 + ϵ, ϵ∼ N(0,1)

Call:

lm(formula = bodyfat.percent ~ AbdomenC + Weight + Height + NeckC +

ChestC + HipC + ThighC + KneeC + AnkleC + BicepsC + ForearmC +

WristC, data = bodyfat)

Residuals:

Min 1Q Median 3Q Max

-14.1320 -3.1309 -0.2276 3.2606 9.2764

Coefficients:

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

(Intercept) 4.68301 23.94904 0.196 0.8451

AbdomenC 1.03512 0.08425 12.286 <2e-16 \*\*\*

Weight -0.05043 0.06748 -0.747 0.4556

Height -0.30016 0.19614 -1.530 0.1273

NeckC -0.36143 0.24045 -1.503 0.1341

ChestC -0.14946 0.11089 -1.348 0.1790

HipC -0.21225 0.14910 -1.423 0.1559

ThighC 0.07923 0.14196 0.558 0.5773

KneeC 0.07019 0.24391 0.288 0.7738

AnkleC 0.23758 0.22444 1.059 0.2909

BicepsC 0.27137 0.17448 1.555 0.1212

ForearmC 0.22654 0.21088 1.074 0.2838

WristC -1.61083 0.51300 -3.140 0.0019 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.369 on 237 degrees of freedom

Multiple R-squared: 0.7399, Adjusted R-squared: 0.7268

F-statistic: 56.19 on 12 and 237 DF, p-value: < 2.2e-16

**Least square estimates of the parameters**:

β̂0=4.68301, β̂1=1.03512, β̂2=- 1.03512, β̂3=- 0.30016, β̂4=- 0.36143, β̂5=- 0.14946, β̂6=- 0.21225, β̂7=0.07923, β̂8=0.07019, β̂9=0.23758, β̂10=0.27137, β̂11=0.22654, β̂12=- 1.61083, σ̂2=4.3692=19.088

**The equation of the estimated regression model**:

Y = 4.68301 + 1.03512X1 - 1.03512X2 - 0.30016X3 - 0.36143X4 - 0.14946X5 - 0.21225X6 + 0.07923X7

+ 0.07019X8 + 0.23758X9 + 0.27137X10 + 0.22654X11 - 1.61083X12

**The value of the determination coefficient**: R2 = 0.7399

**Hypothesis Test**:

H0: β0=β1=β2=β3=β4=β5=β6=β7=β8=β9=β10=β11=β12=0

H1: at least one of the predictors is significantly different from zero.

Test Statistic is F distribution with F12,237 = 56.19

p-value < 2.2 × 10-16

Since p-value < 2.2 × 10-16 < α =0.05, we reject null hypothesis.

Thus, at least one of the predictors is significantly different from zero.

**Is this model better**?

Comparing with the model of Abdomen Circumference, this new model has R2 = 0.7399, which is larger than R2 = 0.6578 of the regression model of Abdomen Circumference.

Also, in this model, we can observe that 2 predictors, Abdomen Circumference and Wrist Circumference, have impact on body fat percentage since their p-value are both smaller than α =0.05.

Thus, this new model is a better model to predict body fat percentage.