STAT8111 Assigment1

Phan Vinh Phu 45747989

2023-08-18

Question 1

a.

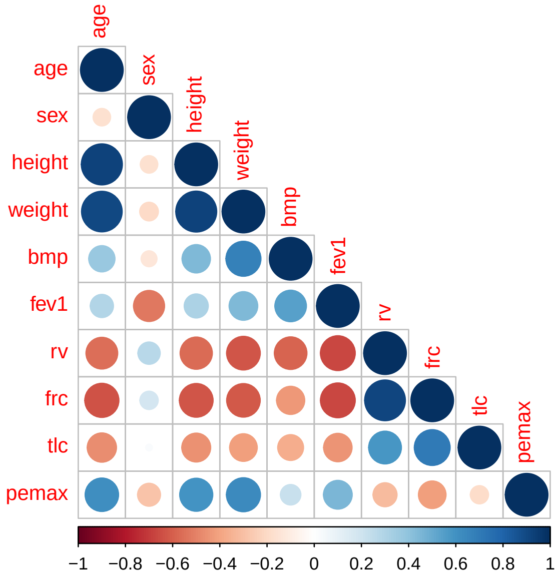
Numerical correlation

```
num_cor <- round(cor(data), 4)
num_cor
```

```
##            age     sex  height  weight     bmp    fev1      rv     frc     tlc
## age     1.0000 -0.1671  0.9261  0.9059  0.3778  0.2945 -0.5519 -0.6394 -0.4694
## sex    -0.1671  1.0000 -0.1675 -0.1904 -0.1376 -0.5283  0.2714  0.1836  0.0242
## height  0.9261 -0.1675  1.0000  0.9207  0.4408  0.3167 -0.5695 -0.6243 -0.4571
## weight  0.9059 -0.1904  0.9207  1.0000  0.6725  0.4488 -0.6215 -0.6173 -0.4185
## bmp     0.3778 -0.1376  0.4408  0.6725  1.0000  0.5455 -0.5824 -0.4344 -0.3649
## fev1    0.2945 -0.5283  0.3167  0.4488  0.5455  1.0000 -0.6659 -0.6651 -0.4430
## rv     -0.5519  0.2714 -0.5695 -0.6215 -0.5824 -0.6659  1.0000  0.9106  0.5891
## frc    -0.6394  0.1836 -0.6243 -0.6173 -0.4344 -0.6651  0.9106  1.0000  0.7044
## tlc    -0.4694  0.0242 -0.4571 -0.4185 -0.3649 -0.4430  0.5891  0.7044  1.0000
## pemax   0.6135 -0.2886  0.5992  0.6352  0.2295  0.4534 -0.3156 -0.4172 -0.1816
##          pemax
## age     0.6135
## sex    -0.2886
## height  0.5992
## weight  0.6352
## bmp     0.2295
## fev1    0.4534
## rv     -0.3156
## frc    -0.4172
## tlc    -0.1816
## pemax   1.0000
```

Graphical correlation

```
corrplot(num_cor, type="lower")
```



Comment :

From the above correlation plot and correlation table, it can be seen that :

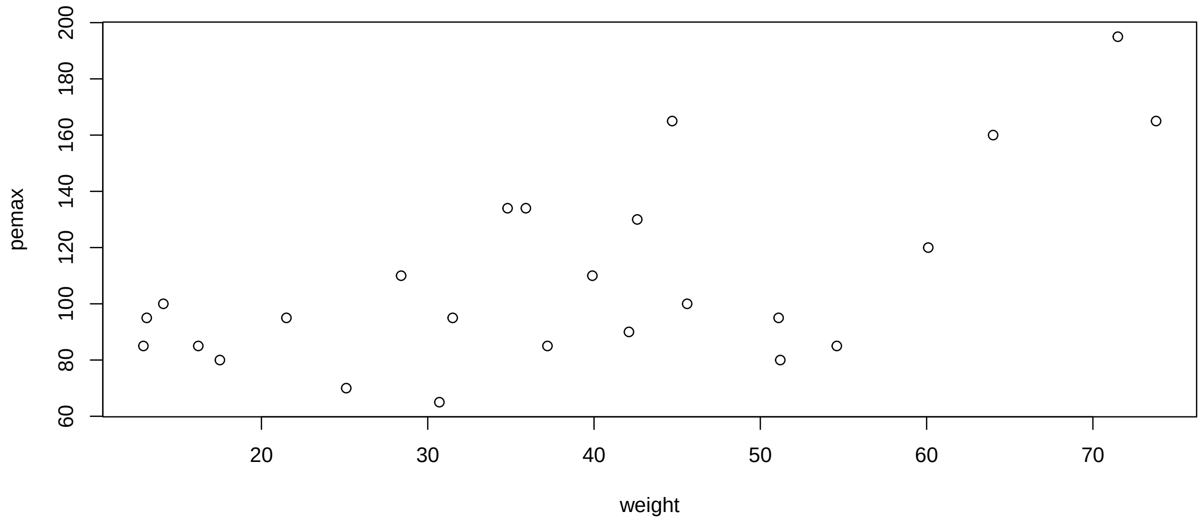
- There might be a strong positive linear relationship between `age` and `height`, `weight` and `pemax`. Beside that, `age` is negatively correlated with `frc`.
- `height` and `weight` are strongly correlated. In addition, `weight` has moderate negative correlation with `rv`, `frc`, `tlc` and positive correlation with `bmp`, `pemax`.
- Similarly, `fev1` has moderate negative correlation with `rv`, `frc`, `tlc`.
- Lastly, `rv` highly positively correlated with `frc`.

b.

In this part, the relationship between `weight` and `pemax` will be examined.Specifically, `pemax` is the dependent variable (Y) and `weight` is the independent variable (X).

Scatter Plot

```
plot(x = data$weight, y = data$pemax, xlab="weight", ylab = "pemax")
```



The graph illustrates that `weight` is positively related with `pemax`. The trend is that when `weight` increase, `pemax` will increase.

Linear Model

```
model <- lm(pemax ~ weight, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = pemax ~ weight, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44.30 -22.69   2.23  15.91  48.41
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.5456    12.7016   5.003 4.63e-05 ***
## weight        1.1867     0.3009   3.944 0.000646 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.38 on 23 degrees of freedom
## Multiple R-squared:  0.4035, Adjusted R-squared:  0.3776
## F-statistic: 15.56 on 1 and 23 DF,  p-value: 0.0006457
```

The model equation $pe\hat{m}ax = 63.5456 + 1.1867 weight + \epsilon \;\; (\epsilon \sim N(0, \sigma^2))$

Model Fit

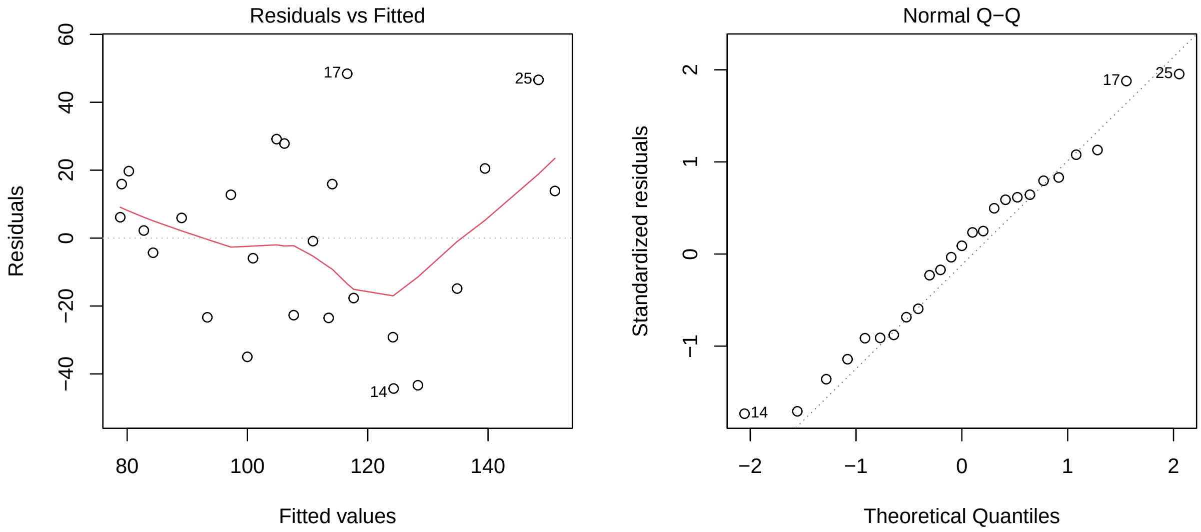
- The $R^2 = 0.4035$ which means 40.35% of the variation in `pemax` can be explained by `weight`. This shows that the model does not fit well.

Model interpretation

- According to the equation, for each unit increase in weight, the `pemax` will increase about 1.1867.
- `weight` is significant predictor since the p-value = 0.000646 (< 0.001).

```
par(mfrow = c(1,2))
plot(model, which = c(1,2))
```

Diagnostic Checking



- The standardized Residuals versus Fitted values plot appears to be a random scatter about zero, so the model is adequate.This graph also show some residuals which are 14, 17, 25 are low and high. This can be evidence of small amount of heteroscedasticity.
- the Normal Q-Q plot is approximately linear, so it can be said that the normality assumption holds.

c.

Model 2

```r
model_2 <- lm(pemax ~ weight + sex, data = data)
summary(model_2)
```

```
##
## Call:
## lm(formula = pemax ~ weight + sex, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.388 -16.850   0.073  13.168  43.748
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  70.9719    14.4644   4.907 6.61e-05 ***
## weight        1.1248     0.3056   3.681  0.00131 **
## sex         -11.4776    10.7963  -1.063  0.29926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.31 on 22 degrees of freedom
## Multiple R-squared:  0.4327, Adjusted R-squared:  0.3811
## F-statistic: 8.388 on 2 and 22 DF,  p-value: 0.00196
```

The model 2 equation $\widehat{pemax} = 70.9719 + 1.1248 weight + (-11.4776) sex_i + \epsilon \quad (\epsilon \sim N(0, \sigma^2))$

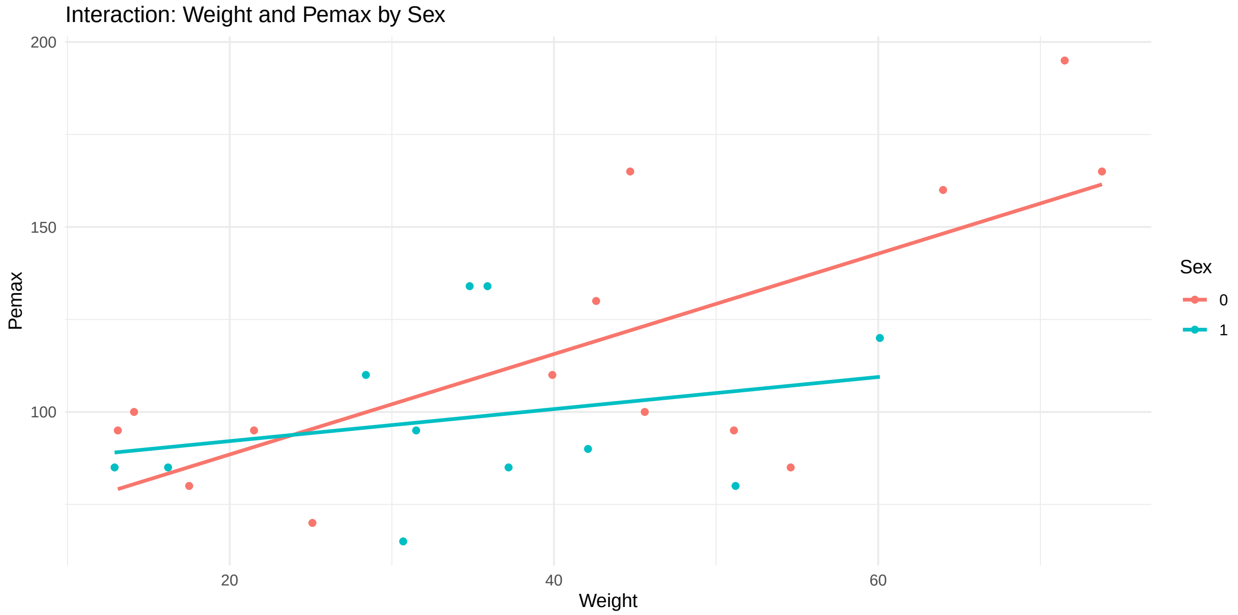
Model 2 analysis

- This model includes the `sex` and `weight` variables as a predictor for `pemax`
- `sex` is insignificant predictor (with p-value = 0.29926) since there is no clear relationship between `sex` and `pemax`.
- `weight` is significant predictor with p-value = 0.00131
- The $R^2 = 0.4327$ indicates that the model does not fit well, only 43.27% of variation in `pemax` can be explained by this model.
- the p-value for the F-statistic is 0.00196, indicating that the model as a whole is statistically significant.

Model 3

Check the interaction plot

```
# Create an interaction plot
ggplot(data, aes(x = weight, y = pemax, color = factor(sex))) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE) +
  labs(title = "Interaction: Weight and Pemax by Sex",
       x = "Weight", y = "Pemax", color = "Sex") +
  theme_minimal()
```



The lines of graph are not parallel which shows that there is a interaction between `weight` and `pemax` with respect to `sex`. This interaction effect indicates that the relationship between `weight` and `pemax` variable depends on the sex of the individual. This suggests that there might be an interaction effect between `weight` and `sex` in predicting `pemax`.

```
model_3 <- lm(pemax ~ weight*sex, data = data)
summary(model_3)
```

```
##
## Call:
## lm(formula = pemax ~ weight * sex, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.464 -14.565  -2.096  14.247  42.973
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   61.3603    15.9335   3.851 0.000927 ***
## weight        1.3572     0.3471   3.910 0.000805 ***
## sex          22.0905    27.2923   0.809 0.427358
## weight:sex   -0.9240     0.6922  -1.335 0.196187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.85 on 21 degrees of freedom
## Multiple R-squared:  0.477,  Adjusted R-squared:  0.4023
## F-statistic: 6.385 on 3 and 21 DF,  p-value: 0.003025
```

The model 3 equation

$\hat{pemax} = 61.3603 + 1.3572 weight + 22.0905 sex + (-0.9240) weight \times sex + \epsilon \quad (\epsilon \sim N(0, \sigma^2))$

Model 3 analysis

- Model 3 includes both `weight` and `sex` as predictors for `pemax` as well as an interaction term between them.
- Only `weight` is the significant predictor with p-value = 0.0008. Beside that, `sex` and the interaction term are insignificant. This means The interaction is insignificant in this case.
- The $R^2 = 0.477$ is higher than in Model 2, indicating that this model explains more variability in the response.
- The F-statistic is significant (p-value: 0.003025), suggesting that the overall model is a better fit than a model without predictors.

Conclusion

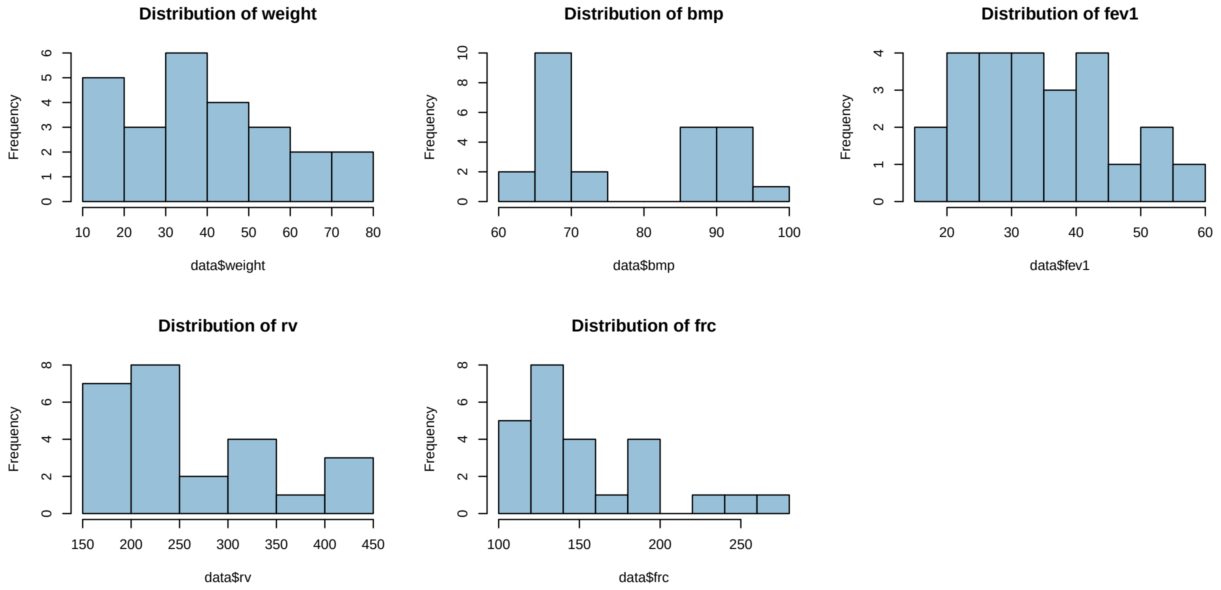
Based on the analysis, Model 3 (including both "weight" and "sex" with interaction) seems to be the best model among three model for predicting "pemax" compared to Model 1 (only `weight`) and Model 2 (including `sex` and `weight`). Model 3 (with $R^2 = 0.477$ and $adjusted R^2 = 0.4023$) provides more insight into the relationship between the predictors and the response, and it demonstrates a better overall fit to the data.

d.

We are interested in predicting `pemax` (maximal expiratory pressure) by using `weight`, `bmp`, `fev1`, `rv`,`frc`. In this case, the stepwise backward selection will be used to find the best model. In stepwise backward selection, people first regress with all predictor variables in the model. Then, the predictors with the largest p-value in the t-test will be dropped. Next step is fitting the reduced model. This progess will be run iteratively until all variables in the model are significant. First, we will check the distributions of independent variables.

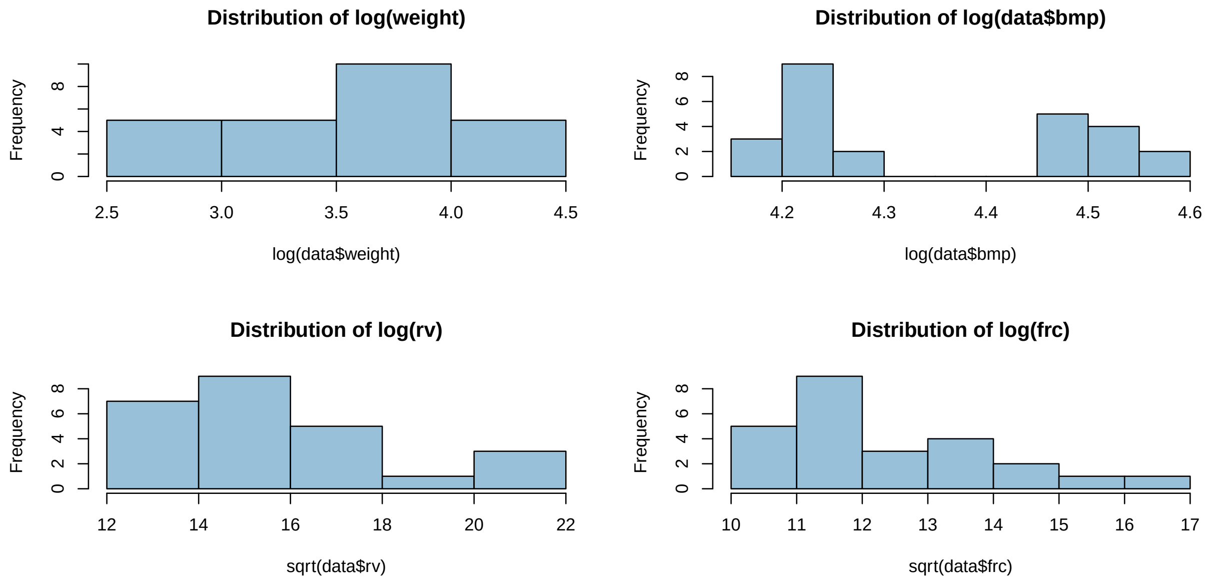
Exam the histogram

```r
# Obtain histograms
par(mfrow = c(2,3))
hist(data$weight,
     main="Distribution of weight",
     col="#98c1d9")
hist(data$bmp,
     main="Distribution of bmp",
     col="#98c1d9")
hist(data$fev1,
     main="Distribution of fev1",
     col="#98c1d9")
hist(data$rv,
     main="Distribution of rv",
     col="#98c1d9")
hist(data$frc,
     main="Distribution of frc",
     col="#98c1d9")
```



The graphs indicate that most of variables are right-skewed except for `fev1`. Therefore, I will try the log tranformation for them.

```
# Obtain histograms after log transformation
par(mfrow = c(2,2))
hist(log(data$weight),
     main="Distribution of log(weight)",
     col="#98c1d9")
hist(log(data$bmp),
     main="Distribution of log(data$bmp)",
     col="#98c1d9")
hist(sqrt(data$rv),
     main="Distribution of log(rv)",
     col="#98c1d9")
hist(sqrt(data$frc),
     main="Distribution of log(frc)",
     col="#98c1d9")
```



According to the histograms,the log transformation is not effective in this case. Hence, the normal form of variables will be use for the linear model. Since I have examined the correlation plot and correlation table in question 1a, I will reuse the above information to advoid the collinearities .Now, Let's start with the full model.

Full Model

```
full_model <- lm(pemax ~ weight + bmp + fev1 + rv + frc , data = data)
summary(full_model)
```

```
## 
## Call:
## lm(formula = pemax ~ weight + bmp + fev1 + rv + frc, data = data)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -39.72 -12.17   4.83  15.29  34.75 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 64.18640   54.73550   1.173 0.255423    
## weight       1.73556    0.42529   4.081 0.000637 ***
## bmp         -1.35105    0.66763  -2.024 0.057303 .  
## fev1         1.53087    0.62948   2.432 0.025078 *  
## rv           0.13612    0.15668   0.869 0.395787    
## frc         -0.02477    0.31278  -0.079 0.937703    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 23.34 on 19 degrees of freedom
## Multiple R-squared:  0.6142, Adjusted R-squared:  0.5127 
## F-statistic:  6.05 on 5 and 19 DF,  p-value: 0.001637
```

It can be seen that frc is insignificant predictor at significant level 5% and has the largest p-value which is 0.937703. Therefore, frc explains the least variation when added to the model. Now , frc will be dropped.

The reduced model without frc

```
reduced_model <- lm(pemax ~ weight + bmp + fev1 + rv , data = data)
summary(reduced_model)
```

```
##
## Call:
## lm(formula = pemax ~ weight + bmp + fev1 + rv, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -39.77 -11.74   4.33  15.66  35.07
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 63.94669   53.27673   1.200 0.244057
## weight       1.74891    0.38063   4.595 0.000175 ***
## bmp         -1.37724    0.56534  -2.436 0.024322 *
## fev1         1.54770    0.57761   2.679 0.014410 *
## rv           0.12572    0.08315   1.512 0.146178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.75 on 20 degrees of freedom
## Multiple R-squared:  0.6141, Adjusted R-squared:  0.5369
## F-statistic: 7.957 on 4 and 20 DF,  p-value: 0.000523
```

Similarly, rv is the insignificant predictor at significant level 5% and has the largest P-value which is 0.146178. Hence, I drop rv and fit the reduced model without frc and rv.

The reduced model without frc and rv

```
reduced_model_2 <- lm(pemax ~ weight + bmp + fev1 , data = data)
summary(reduced_model_2)
```

```
##
## Call:
## lm(formula = pemax ~ weight + bmp + fev1, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.388 -13.496   3.991  14.856  40.373
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 126.3336    34.7199   3.639 0.001536 **
## weight        1.5365     0.3644   4.216 0.000387 ***
## bmp          -1.4654     0.5793  -2.530 0.019486 *
## fev1          1.1086     0.5144   2.155 0.042893 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.44 on 21 degrees of freedom
## Multiple R-squared:   0.57,  Adjusted R-squared:  0.5086
## F-statistic: 9.279 on 3 and 21 DF,  p-value: 0.000418
```

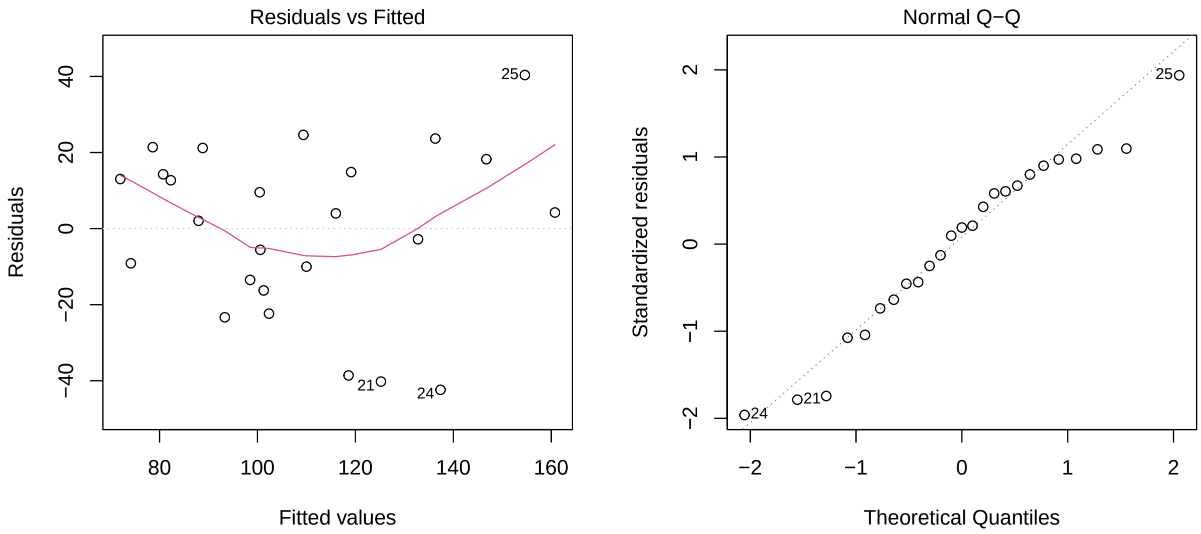
```
# Calculate the correlation matrix
correlation_matrix <- cor(data[, c("weight", "bmp", "fev1")])

# Print the correlation matrix
print(correlation_matrix)
```

```
##           weight       bmp      fev1
## weight 1.0000000 0.6725463 0.4488393
## bmp    0.6725463 1.0000000 0.5455204
## fev1   0.4488393 0.5455204 1.0000000
```

At this stage, all predictors are significant, therefore, selection process stops here. Nevertheless, there is moderate collinearities since `bmp`, `weight`, `fev1` are positively correlated. Their correlation are insignificant because all correlation scores are less than 0.7.

Diagnostic Checking

```
par(mfrow = c(1,2))
plot(reduced_model_2, which = 1:2)
```



- The Residuals vs Fitted plot look like random scatter around 0. Therefore, there is no obvious pattern in any of the residual plots so it appears the linearity and constant variance assumptions of the multiple linear model are justified.
- The quantile plot of residuals look approximately linear, suggesting the normality assumption for residuals is appropriate.

The final model equation

$\hat{pemax} = 126.3336 + 1.5365 weight + (-1.4654) bmp + 1.1086 fev1 + \epsilon \quad (\epsilon \sim N(0, \sigma^2))$

Interpretaion

From above equation,

- For every unit increase in `weight` (`bmp` and `fev1` stay unchanged), the average `pemax` is expected to increase by 1.5365 units.
- For every unit increase in `bmp` (holding `weight` and `fev1` constant), the average `pemax` is expected to decrease by 1.4654 units.
- For one extra unit of `fev1` , the average `pemax` is expected to increase by 1.1086 units.
- The R-squared value of 0.57 indicates that approximately 57% of the variability in the `pemax` can be explained by the predictor variables (`weight`, `bmp`, and `fev1`) included in the model.
- The final mode has a balance between model complexity and goodness of fit.

Question 2

a.

The Inverse Gaussian distribution is defined a

$$f(x; \mu, \gamma) = \sqrt{\frac{\gamma}{2\pi x^3}} \exp\left(-\frac{\gamma (x-\mu)^2}{2\mu^2 x}\right)$$

$$f(x; \mu, \gamma) = \exp\left(\frac{1}{2}\log\left(\frac{\gamma}{2\pi x^3}\right) - \frac{\gamma}{2\mu^2}\frac{(x-\mu)^2}{x}\right)$$

$$f(x; \mu, \gamma) = \exp\left(\frac{1}{2}\log\left(\frac{\gamma}{2\pi x^3}\right) - \frac{\gamma}{2\mu^2}\frac{(x^2 - 2x\mu + \mu^2)}{x}\right)$$

$$f(x; \mu, \gamma) = \exp\left(\frac{1}{2}\log\left(\frac{\gamma}{2\pi x^3}\right) - \frac{\gamma x}{2\mu^2} + \frac{x}{\mu} + \frac{\gamma}{2x}\right)$$

$$f(x; \mu, \gamma) = \exp\left(-\frac{\gamma x}{2\mu^2} + \frac{x}{\mu} + \frac{1}{2}\log\left(\frac{\gamma}{2\pi x^3}\right) + \frac{\gamma}{2x}\right)$$

The simpler form of the pdf:

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

in this case :

$$\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi) = -\frac{\gamma x}{2\mu^2} + \frac{x}{\mu} + \frac{1}{2}log(\frac{\gamma}{2\pi x^3}) + \frac{\gamma}{2x}$$

with:

$$\phi = \frac{1}{\gamma},\ a(\phi) = \frac{1}{\gamma},\ \theta = -\frac{1}{2\mu^2},\ b(\theta) = 2\theta^{\frac{1}{2}}, c = \frac{1}{2}log(\frac{\theta}{2\pi x^3}) + \frac{\theta}{2x}$$

.

b.

In the context of the exponential family representation of the Inverse Gaussian distribution, the natural parameter and the scale parameter can be identified as follows :

Natural Parameter is $-\frac{\gamma}{2\mu^2}$.

Scale Parameter is μ.

c.

Mean

$$b'(\theta) = \frac{1}{\sqrt{2\theta}}, \quad \text{where } \theta = -\frac{1}{2\mu^2}.$$

$$\text{Substitute } \theta = -\frac{1}{2\mu^2} :$$

$$b'(\theta) = \frac{1}{\sqrt{2\left(-\frac{1}{2\mu^2}\right)}}$$

Simplify:

$$b'(\theta) = \frac{1}{\sqrt{-\frac{1}{\mu^2}}} = \frac{1}{\frac{1}{\mu}} = \mu$$

So, $E[X] = \mu$

Variance

$$Var(Y) = b''(\theta)\phi$$

$$Var(Y) = \phi\theta^{-\frac{3}{2}}$$

$$Recall\ that\ \theta^{-\frac{1}{2}} = \mu\ and\ \phi = \frac{1}{\gamma}$$

$$\Rightarrow\ Var(Y) = \frac{\mu^3}{\gamma}$$

Question 3

The normal linear model is

$$Y_i = x_i^T \beta + \epsilon_i \quad \text{for } i = 1, \dots, n$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ is the p-dimensional regressor parameter, and the ϵ_i are the noise of the model.

a.

Given the Normal distribution for the noise ϵ_i as $N(0, \sigma^2)$, the likelihood of y_i is:

$$L_i(\beta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i^T\beta}{\sigma}\right)^2\right)$$

The likelihood of the entire sample y is the product of the individual likelihoods:

$$L(\beta) = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i^T\beta}{\sigma}\right)^2\right)$$

Now, let's consider the log-likelihood:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{y_i - x_i^T\beta}{\sigma}\right)^2 \\ &= -n\log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 \end{aligned}$$

To find the maximum likelihood estimator of β, we want to maximize $l(\beta)$ with respect to β. Since maximizing the log-likelihood is equivalent to maximizing the likelihood itself, we need to find the value of β that maximizes the expression above.

Maximizing $l(\beta)$ with respect to β is equivalent to minimizing the following expression, which is proportional to the squared error loss:

$$LS(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2$$

Therefore, the maximum likelihood estimator of the parameter β is also the solution of the squared error loss, known as the least square error (LS), given by $LS(\beta)$. This demonstrates the connection between the maximum likelihood estimation and the least squares estimation.

b.

Given the Laplace distribution for the noise ϵ_i as $L(0, \sigma^2)$, the likelihood of y_i is:

$$L_i(\beta) = \frac{1}{2\sigma} \exp\left(-\frac{|y_i - x_i^T\beta|}{\sigma}\right)$$

The likelihood of the entire sample y is the product of the individual likelihoods:

$$L(\beta) = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} \frac{1}{2\sigma} \exp\left(-\frac{|y_i - x_i^T\beta|}{\sigma}\right)$$

Now, let's consider the log-likelihood:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \log\left(\frac{1}{2\sigma}\right) - \frac{|y_i - x_i^T\beta|}{\sigma} \\ &= -n\log(2\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n}|y_i - x_i^T\beta| \end{aligned}$$

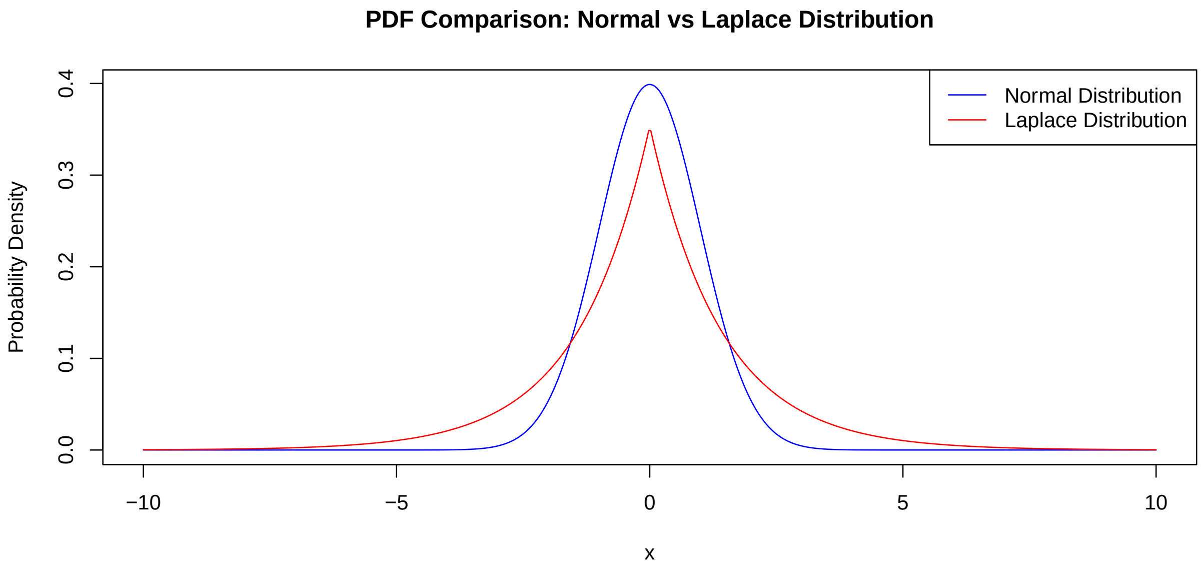
To maximize $l(\beta)$ with respect to β, we want to maximize the negative of the absolute error loss, which is equivalent to minimizing the absolute error loss itself:

$$AL(\beta) = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i^T\beta|$$

Therefore, in the case of a Laplace distribution error noise, maximizing the log-likelihood is equivalent to minimizing the absolute error loss (AL) defined by $AL(\beta)$. This demonstrates the connection between the maximum likelihood estimation and the minimization of the absolute error loss for the Laplace distribution case.

c.

```
# Parameters
mu <- 0         # Mean for both distributions
sigma <- 1      # Standard deviation for the Normal distribution
scale <- sqrt(2) * sigma  # Scale parameter for Laplace distribution

# Values for x-axis
x <- seq(-10, 10, length.out = 500)

# Calculate pdf values for Normal and Laplace distributions
pdf_normal <- dnorm(x, mean = mu, sd = sigma)
pdf_laplace <- 1 / (2 * scale) * exp(-abs(x - mu) / scale)

# Plotting
plot(x, pdf_normal, type = "l", col = "blue", xlab = "x", ylab = "Probability Density",
     main = "PDF Comparison: Normal vs Laplace Distribution")
lines(x, pdf_laplace, col = "red")
legend("topright", legend = c("Normal Distribution", "Laplace Distribution"),
       col = c("blue", "red"), lty = 1)
```



d.

The claim that the linear model using Laplace error noise is more robust to outliers can be supported by observing the behavior of the probability density functions (pdfs) of the Normal and Laplace distributions in the previous plot.

Here are some arguments based on the plot to support the claim:

- **Heavier Tails of the Laplace Distribution** : In the plot, it can be seen that the Laplace distribution's pdf has heavier tails compared to the Normal distribution. This means that the Laplace distribution assigns more probability to extreme values (outliers) compared to the Normal distribution.
- **Reduced Influence of Outliers** : In the Laplace distribution, the tail behavior implies that outliers will have a smaller impact on the model estimation compared to the Normal distribution. Outliers that are far from the mean will contribute less to the overall loss, making the model less sensitive to extreme observations. On the other hand, the Normal distribution's lighter tails make it more sensitive to outliers.
- **Minimization of Absolute Errors** : The Laplace distribution's density has a sharp peak at the mean, resulting in a higher concentration of values around the center. When you minimize the absolute error loss, which the Laplace distribution corresponds to, you effectively prioritize minimizing the impact of large individual errors (outliers) rather than the sum of squared errors as in the Normal distribution case. This aligns well with the robustness against individual extreme observations.