SM3_Project

Shingyan Kwong

May 13, 2018

Part A.

1.

Here, we produce pairwise scatterplots for height, weight and length to see the individual relationships between these varaibles (Figure 1):

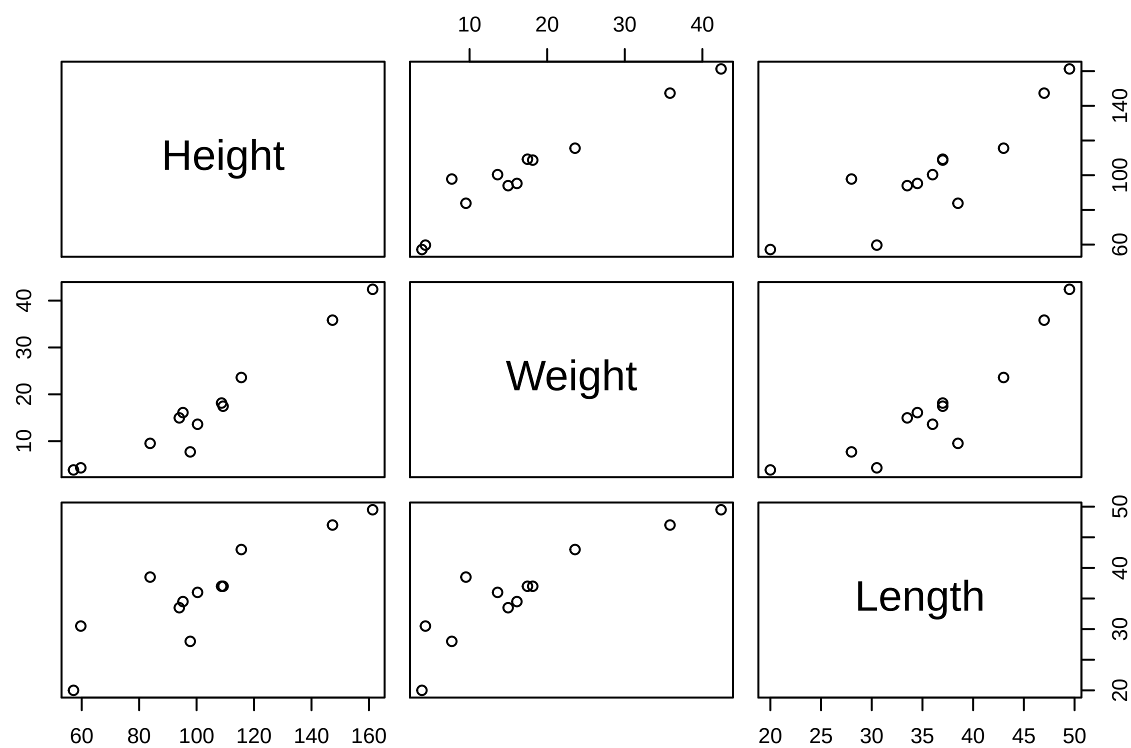

Figure 1: Pairwise Scatterplots for Height, Weight and Length

In figure 1 we see that there is evidence of a strong, positive linear relationship between length and the two predictor variables, height and weight. The associated correlation coefficients are 0.881 and 0.894 respectively. There is also a strong, positive linear relationship between height and weight. This suggests that the two predictors may be dependent one another.

2.

Here we fit three linear models: lm1 which fits Length against Height and Weight; lm2 which fits Length against Height; and lm3 which fits Length against Weight.

```
lm1<-lm(Length~Height+Weight, data=child)
lm2<-lm(Length~Height, data=child)
lm3<-lm(Length~Weight, data=child)
```

3.

The model assumptions which may be checked via diagnostic plots are as follows.

Linearity: Check the residuals vs fitted and the residuals vs predictor plots. Linearity is reasonable if random scatter above and below the 0 line is observed.

Constant Variance: Check scale location plot. Homoscedacity is reasonable if constant variance of residuals is observed across the scale location plot.

Normality: Check normal quantile-quantile plot. Normality is reasonable if most points between -2 and 2 are on/close to the diagonal line.

4.

Full model (lm1)

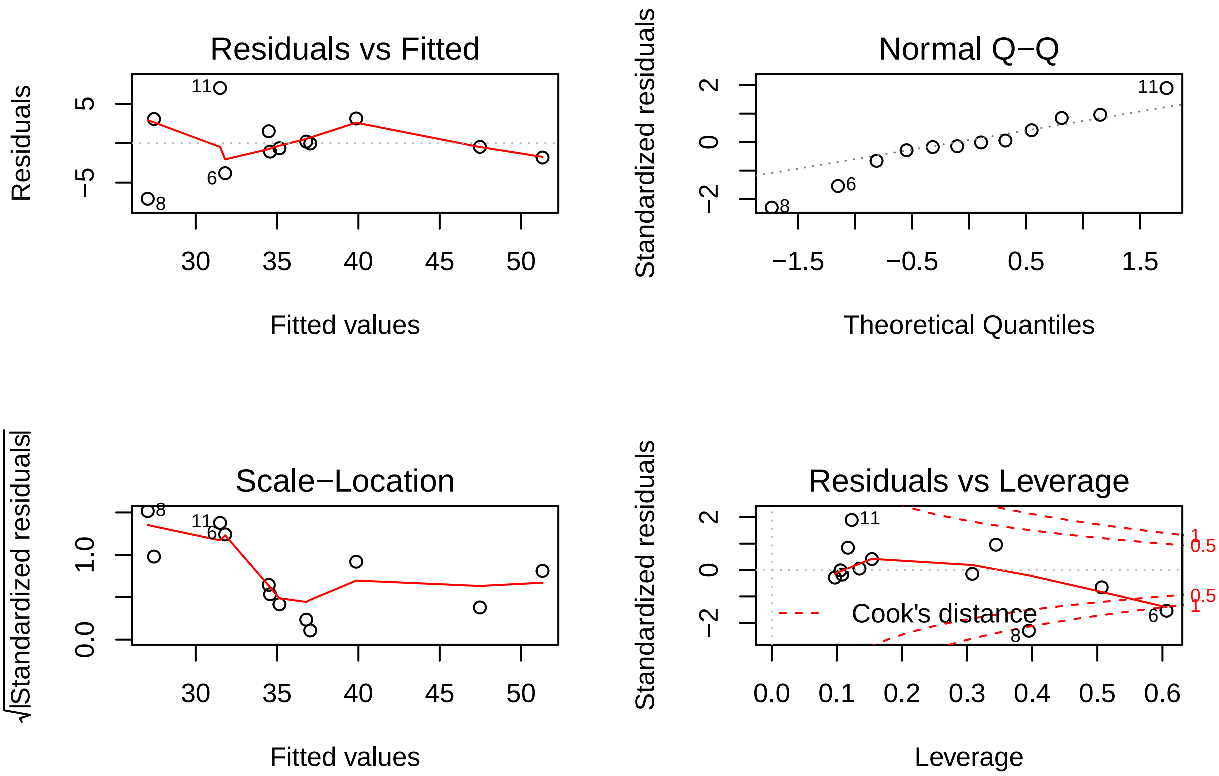

Figure 2: Residuals plots for Full Model

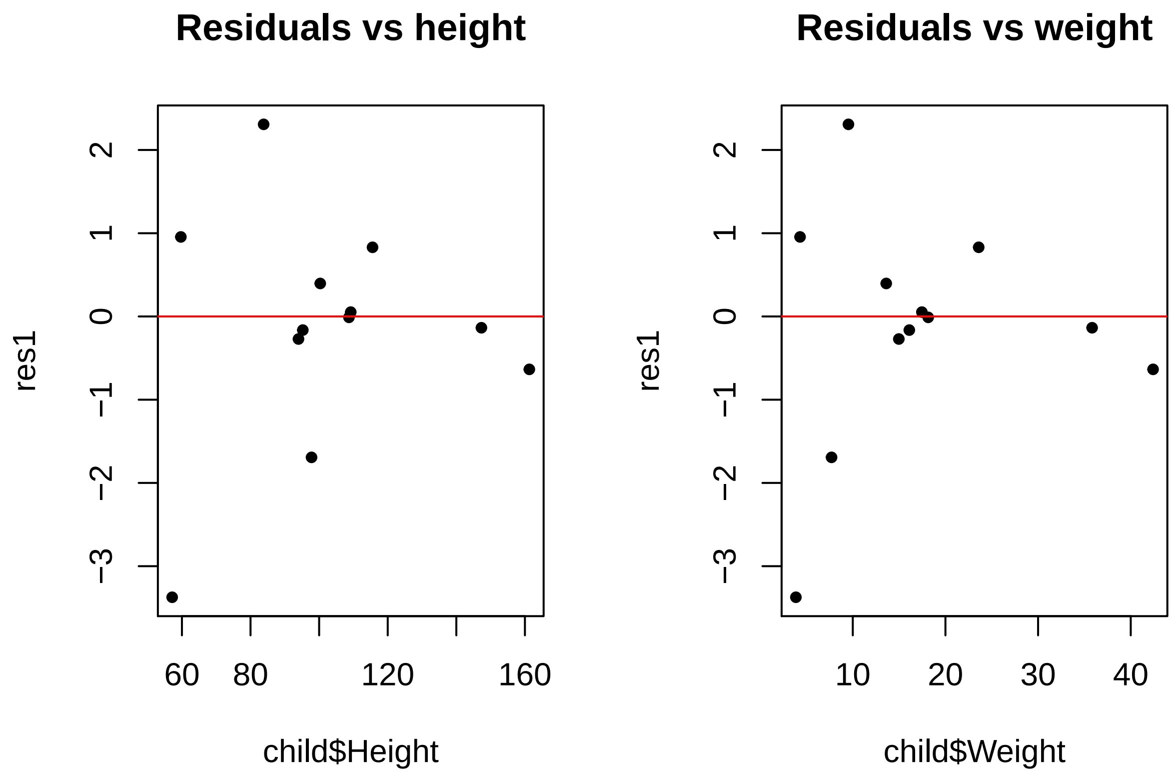

Figure 3: Residuals plots for Height and Weight Individually (Full Model)

From figures 2 and 3 we can make the following assessments:

Linearity: Given the small number of data points available, roughly random scatter is observed in the residual vs fitted and residual vs predictor plots. There is a couple of high residual points but it is not too bad. Linearity is reasonable.

Constant variance: Scale location plots appear to show heteroscedacity. Constant variance is not reasonable.

Normality: There is some minor departure from normality in the beginning and the tails of the standardized residuals. Overall the points are fairly close to the diagonal line. Normality is reasonable.

Leverage: There are 2 data points in the zone of danger. These high leverage points are having a disproportionate effect on the model.

Height Model (lm2)

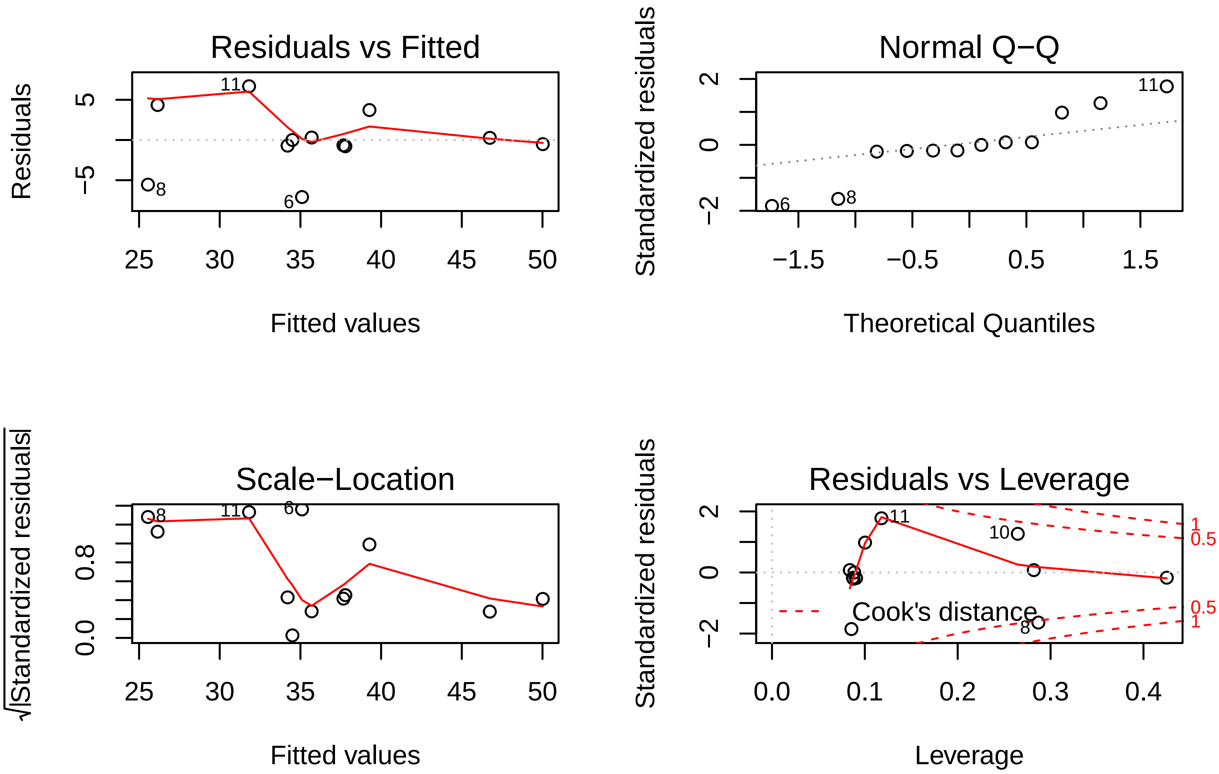

Figure 4: Residuals plots for Height Model

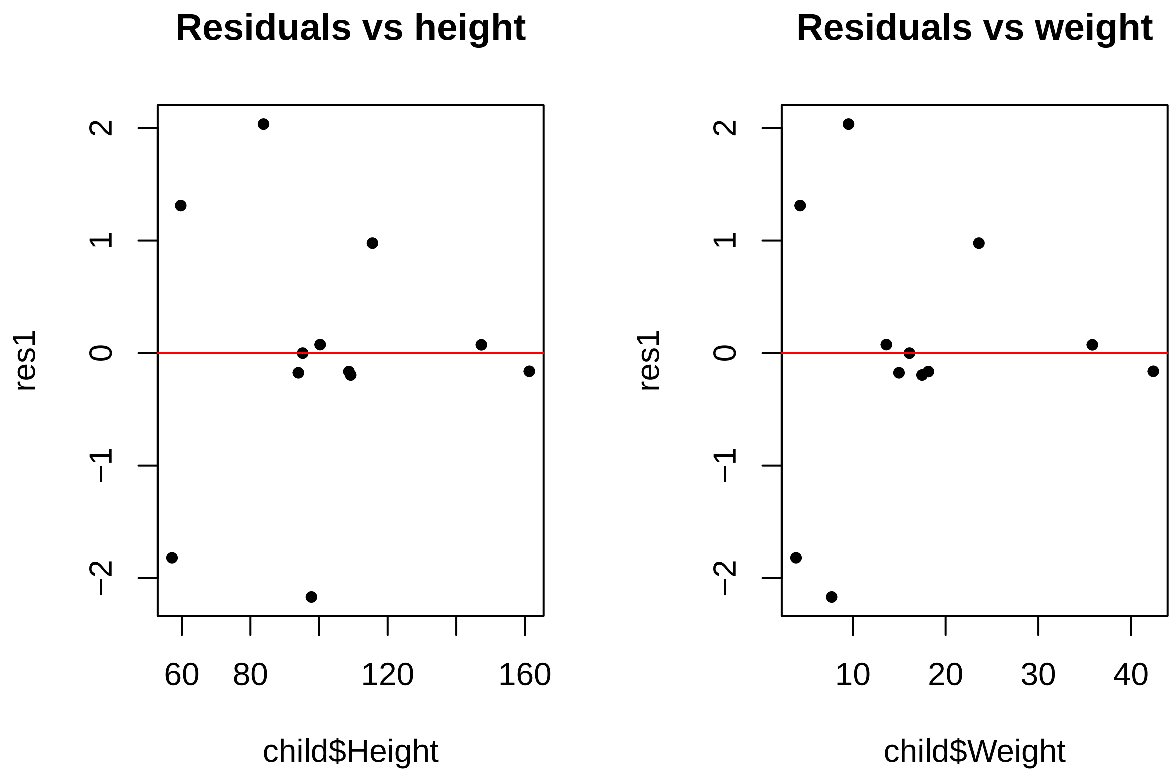

Figure 5: Residuals plots for Height and Weight Individually (Height Model)

From figures 4 and 5 we can make the following assessments:

Linearity: Non-random scatter observed in residual vs fitted and residual vs predictor plots. Linearity is not reasonable.

Constant Variance: Variance appears to increase for the middle fitted values and then decrease again. Constant variance is not reasonable.

Normality: There are several points deviating from the diagonal line on the normal quantile-quantile plot. Normality is not reasonable.

Leverage: There is one point with high leverage.

Weight Model (lm3)

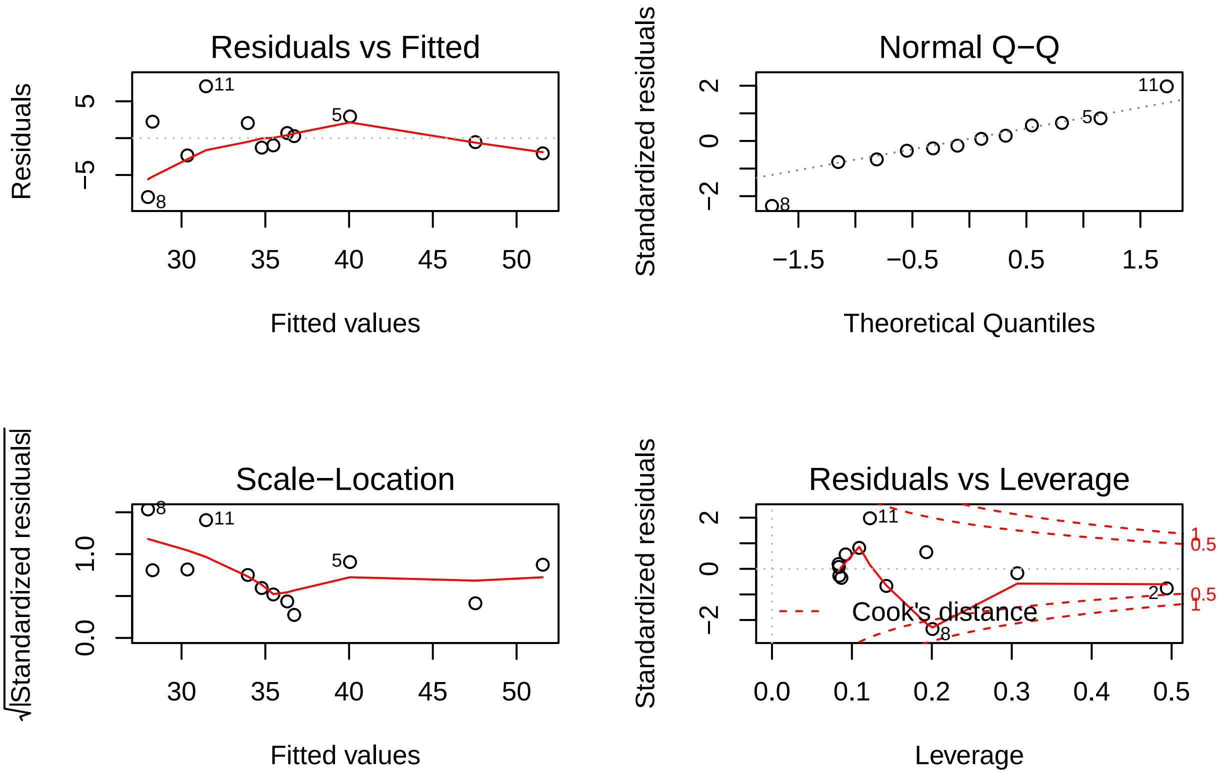

Figure 6: Residuals plots for Weight Model

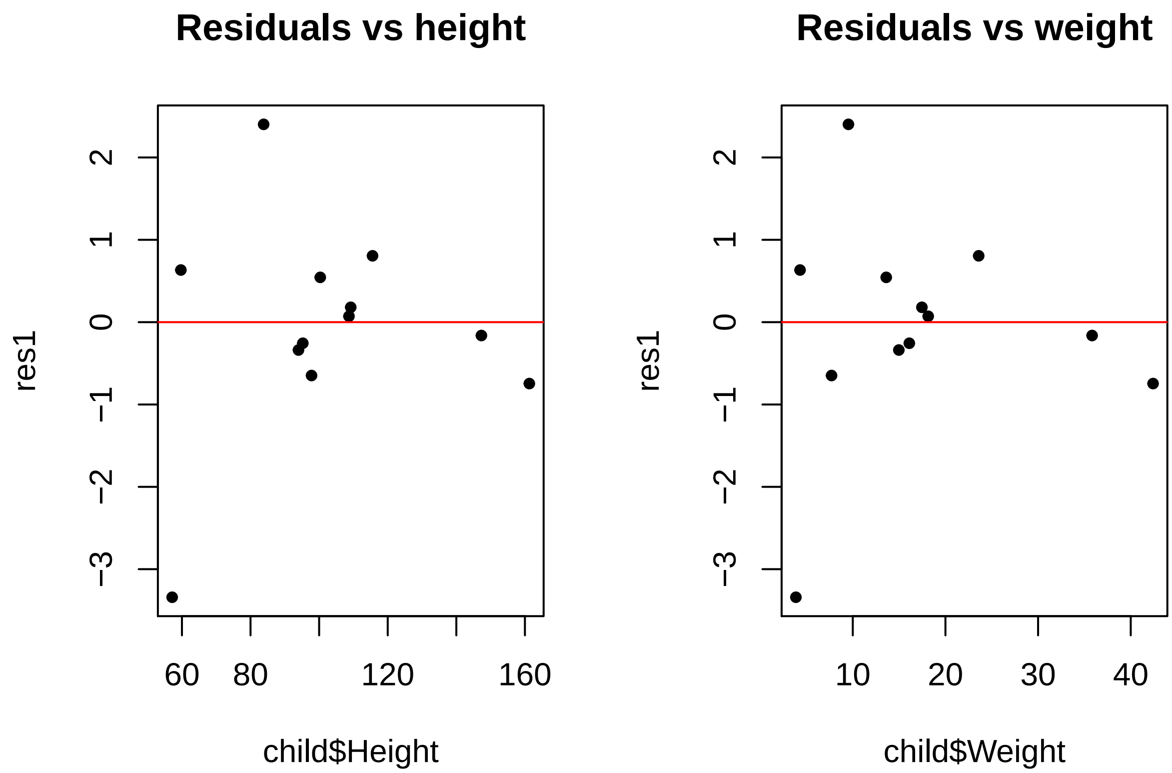

Figure 7: Residuals plots for Height and Weight Individually (Weight Model)

From figure 6 and 7 we can make the following assessments:

Linearity: Rough random scatter in observed in residual vs fitted and residual vs predictor plots. The fitted plot shows some evidence of curvature but overall it is acceptable. Linearity is reasonable.

Constant Variance: Variance is roughly constant across the scale location plot. Constant variance is reasonable.

Normality: Most points are close to the diagonal line except 2. Normality is reasonable.

Leverage: There is one data point with high leverage.

5.

Here we make a comparison of the three models.

We first obtain summaries for each of the three models:

Full Model:

```
##
## Call:
## lm(formula = Length ~ Height + Weight, data = child)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0497 -1.2588 -0.2576  1.8987  7.0030
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.00828    8.74782   2.402   0.0398 *
## Height       0.07729    0.14192   0.545   0.5993
## Weight       0.42081    0.36405   1.156   0.2775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.943 on 9 degrees of freedom
## Multiple R-squared:  0.8054, Adjusted R-squared:  0.7621
## F-statistic: 18.62 on 2 and 9 DF,  p-value: 0.0006332
```

Height Model:

```
##
## Call:
## lm(formula = Length ~ Height, data = child)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0996 -0.7246 -0.2608  1.1585  6.6826
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.12402    4.24711   2.855 0.017113 *
## Height       0.23495    0.03986   5.894 0.000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.008 on 10 degrees of freedom
## Multiple R-squared:  0.7765, Adjusted R-squared:  0.7541
## F-statistic: 34.74 on 1 and 10 DF,  p-value: 0.0001523
```

Weight Model:

```
##
## Call:
## lm(formula = Length ~ Weight, data = child)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9958 -1.4818 -0.1334  2.0899  7.0378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 25.63596    2.00425  12.791 1.60e-07 ***
## Weight       0.61136    0.09698   6.304 8.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.801 on 10 degrees of freedom
## Multiple R-squared:  0.7989, Adjusted R-squared:  0.7788
## F-statistic: 39.74 on 1 and 10 DF,  p-value: 8.865e-05
```

(a)

In the full model, neither predictor variables is statistically significant (at the 0.05 level), and the numerical values of the two coefficients are both smaller than those of the single predictor models.

(b)

Full model:

Holding height constant, the full model predicts that an increase of of 1kg will on average increase the length of the cathetar by 0.42081cm.

Weight only model:

Without regard for height, this model predicts that an increase of 1kg will on average increase the cathetar length by 0.61136cm.

6

(a)

First, we construct the model matrices for the height only and weight only models.

Model Matrix for lm2:

```
##    (Intercept) Height
## 1            1 108.70
## 2            1 161.29
## 3            1  95.25
## 4            1 100.33
## 5            1 115.57
## 6            1  97.79
## 7            1 109.22
## 8            1  57.15
## 9            1  93.98
## 10           1  59.69
## 11           1  83.82
## 12           1 147.32
## attr(,"assign")
## [1] 0 1
```

Model matrix for lm3:

```
##    (Intercept) Weight
## 1            1  18.14
## 2            1  42.41
## 3            1  16.10
## 4            1  13.61
## 5            1  23.59
## 6            1   7.71
## 7            1  17.46
## 8            1   3.86
## 9            1  14.97
## 10           1   4.31
## 11           1   9.53
## 12           1  35.83
## attr(,"assign")
## [1] 0 1
```

Here, we define $\mathbf{1} := (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$, the vector of intercepts for both models. We also denote the vector of height values by $\mathbf{x}_1$ and the vector of weight values by $\mathbf{x}_2$. Then we find that:

$\mathcal{L}_1$ is the space spanned by the columns of M2, that is $\mathcal{L}_1 = span\{\mathbf{1}, \mathbf{x}_1\}$

$\mathcal{L}_2$ is the space spanned by the columns of M3, that is $\mathcal{L}_2 = span\{\mathbf{1}, \mathbf{x}_2\}$

Then, the intersection of the two subspaces is the intercept column, that is $\mathcal{L}_1 \cap \mathcal{L}_2 = span\{\mathbf{1}\}$.

(b)

We note that $(\mathcal{L}_1 \cap \mathcal{L}_2)^\perp$ is the subspace of all vectors orthogonal to $\mathbf{1}$. Then, in order to find the intersections of $\mathcal{L}_1$ and $\mathcal{L}_2$ with $(\mathcal{L}_1 \cap \mathcal{L}_2)^\perp$, we first find orthonormal bases for $\mathcal{L}_1$ and $\mathcal{L}_2$.

We achive this by applying the Gram-Schmidt process. First, we define the basis vectors for both subspaces, and a function norm_vec to find the norm of a vector:

```
one <- c(1,1,1,1,1,1,1,1,1,1,1,1) # intercept vector
x1 <- M2[,2] # vector of height values
x2 <- M3[,2] # vector of weight values
norm_vec <- function(x) sqrt(as.numeric(t(x) %*% x))
```

Next, we find an orthonormal basis for $\mathcal{L}_1$:

```
v1 <- one / norm_vec(one)
v2_ <- x1 - as.numeric((t(x1) %*% v1)) * v1
v2 <- v2_ / norm_vec(v2_)
```

Now

$$\mathbf{v}_1 = \frac{1}{\sqrt{12}}(1,1,1,1,1,1,1,1,1,1,1,1);$$

$\mathbf{v}_2 = (0.0616, 0.5846, -0.0722, -0.0217, 0.1299, -0.0469, 0.0667, -0.4511, -0.0848, -0.4259, -0.1859, 0.4457)$

Then $\mathcal{L}_1 = span\{\mathbf{v}1, \mathbf{v}2\}$.

Now, an orthonormal basis for $\mathcal{L}_2$:

```
w1 <- one / norm_vec(one)
w2_ <- x2 - as.numeric((t(x2) %*% w1)) * w1
w2 <- w2_ / norm_vec(w2_)
```

Now

$$\mathbf{w}_1 = \frac{1}{\sqrt{12}}(1,1,1,1,1,1,1,1,1,1,1,1);$$

$\mathbf{w}_2 = (0.0216, 0.6408, -0.0304, -0.0940, 0.1606, -0.2445, 0.0043, -0.3427, -0.0593, -0.3312, -0.1981, 0.4729)$

Then $\mathcal{L}_2 = span\{\mathbf{w}1, \mathbf{w}2\}$.

Now, we have that $\mathbf{v}_1 = \mathbf{w}_1$ is parallel to $\mathbf{1}$, and that $\mathbf{v}_2$ and $\mathbf{w}_2$ are orthogonal to $\mathbf{1}$, that is, $\mathbf{v}_2, \mathbf{w}_2 \in (\mathcal{L}_1 \cap \mathcal{L}_2)^{\perp}$.

As a result, we find that:

$$\mathcal{L}_1 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^{\perp} = span\{\mathbf{v}_2\};$$

$$\mathcal{L}_2 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^{\perp} = span\{\mathbf{w}_2\}.$$

(c)

Given that both $\mathcal{L}_1 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^{\perp}$ and $\mathcal{L}_2 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^{\perp}$ are one-dimensional subspace, we can compute the angle between them using the relation:

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}|| \; ||\mathbf{v}||}$$

Where $\mathbf{u}$ and $\mathbf{v}$ are two vectors and θ is the angle between them.

We compute the angle between the two spaces in **(b)** as follows:

```
inner <- t(v2) %*% (w2) #Defining the inner product of v2 and w2
norm_v2 <- norm_vec(v2)
norm_w2 <- norm_vec(w2)
theta <- acos(inner/(norm_v2 * norm_w2)) #Finding the angle between v2 and w2
```

Now we have that the angle between $\mathbf{v}_2$ and$\mathbf{w}_2$ is given by:

```
##           [,1]
## [1,] 0.2799352
```

The angle is not π indicating that the two spaces are not orthogonal. This suggests that height and weight are not independent. In fact they are fairly correlated.

7.

In this section we decide upon a final model for the data. We note that since the subspaces $\mathcal{L}_1 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^\perp$ and $\mathcal{L}_2 \cap (\mathcal{L}_1 \cap \mathcal{L}_2)^\perp$ are not orthogonal, the weight and height predictors are not independent; more specifially, they have a significant correlation of 0.961. As such, the full model in which Length was fitted against both Height and Weight should not be used. Now, whjat remains is to decide between the individual height and weight models. In figure 4, we see that the assumptions of linearity, constant variance and normality fail in the height model, and in figure 6 it is clear that these assumptions are much more reasonanble in the weight model. As a result, we can conclude that the most appropriate model for the data is lm3, the model in which length is fit against height alone.

Part B

Introduction

In this section we obtain a predictive model for mammographic mass severity, a measure of the status of mammographic mass lesions, on a scale from 0 to 1, where 0 is assigned to a benign tumor, and 1 is assigned to a malignant tumor. Interest in this analysis arises from there being a low predicitve value of breast biopsy from mammograms. This low predictive value has been found to lead to approximately 70% of unnessessary biopsies of benign tumors. Analysis is performed on the dataset “mammo”, containing the true status of 961 mammographic mass lesions, with the response variable severity as described. Four response variables are considered:

Age - the patient’s age in years;

Shape - a factor variable with four levels: 1 for round, 2 for oval, 3 for lobular, and 4 for irregular;

Margin - a factor varaible with five levels: 1 for circumscribed, 2 for microlobulated, 3 for obscured, 4 for ill-defined, and 5 for spiculated;

Density - a factor with four levels: 1 for high, 2 for iso, 3 for low, and 4 for fat-containing.

The aim of this study is to produce the best predictive model for mammogram mass severity. The model must be parsimonious whilst having minimun residual error. Predicted probabilities of severity may also be generated from the model. This information may be used by clinicians to determine whether a biopsy is appropriate for the patient.

This introduction should probably be reworked but I this hope is a good starting point

Data Entry and Cleaning

First, we enter the data and define any values which are assigned question marks to be missing values.

We then note that BI.RADS is not a predictor variable, and remove it from our analysis:

We generate a correlation matrix to observe any relationships between the predictor variables and the response. It was necessary to omit observations with NA values. Shape and Margin were included as increases in their indices are associated with a greater risk of cancer. Density was omitted as the index is not associated with a greater risk. Thus correlation between Density and Severity would not be statistically meaningful.

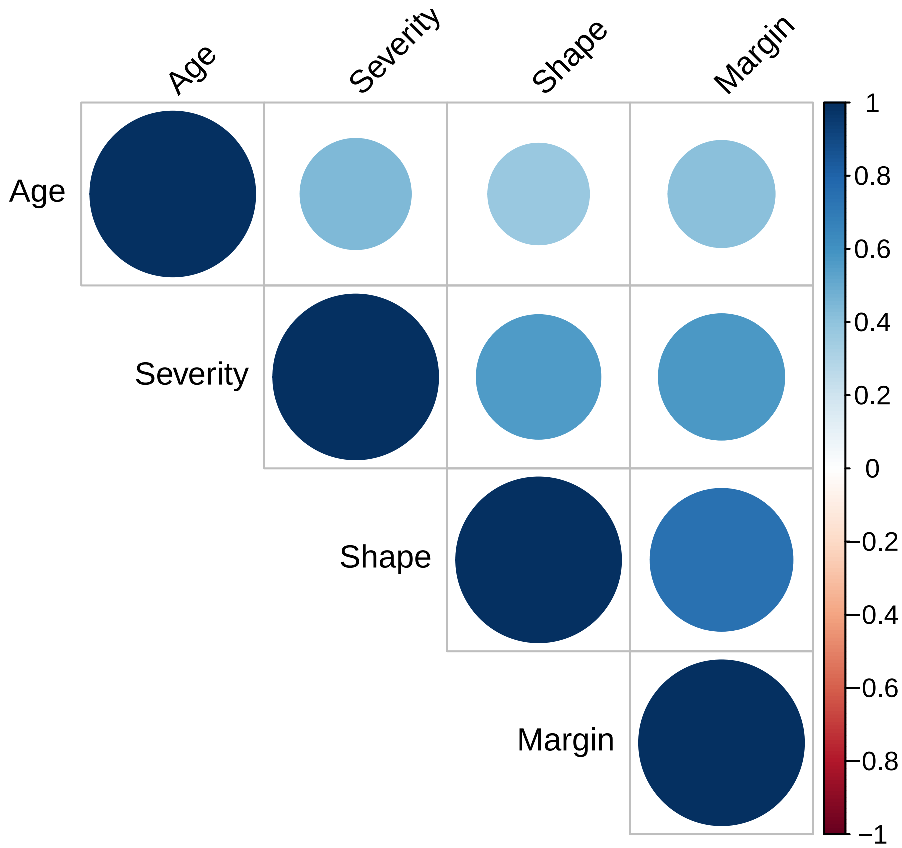

Figure 8: Correlation matrix depicting the relationship between Margin, Shape, Severity and Age.

Mild positive correlation is observed between Age and the other variables.

Moderate positive correlation is observed between Severity and the other 3 predictors.

Strong positive correlation between Shape and Margin.

We can now check the variable types for the data:

```
## 'data.frame':    961 obs. of  5 variables:
##  $ Age     : int  67 43 58 28 74 65 70 42 57 60 ...
##  $ Shape   : int  3 1 4 1 1 1 NA 1 1 NA ...
##  $ Margin  : int  5 1 5 1 5 NA NA NA 5 5 ...
##  $ Density : int  3 NA 3 3 NA 3 3 3 3 1 ...
##  $ Severity: int  1 1 1 0 1 0 0 0 1 1 ...
```

We note that Shape, Margin, Density and Severity should all be factor variables, and as such convert them:

We now see that all of the data types are correct:

```
## 'data.frame':    961 obs. of  5 variables:
##  $ Age     : int  67 43 58 28 74 65 70 42 57 60 ...
##  $ Shape   : Factor w/ 4 levels "1","2","3","4": 3 1 4 1 1 1 NA 1 1 NA ...
##  $ Margin  : Factor w/ 5 levels "1","2","3","4",..: 5 1 5 1 5 NA NA NA 5 5 ...
##  $ Density : Factor w/ 4 levels "1","2","3","4": 3 NA 3 3 NA 3 3 3 3 1 ...
##  $ Severity: Factor w/ 2 levels "0","1": 2 2 2 1 2 1 1 1 2 2 ...
```

Data Visualisations and Data Summaries

To visualise the data, we first produce summary statisitics for the dataset as a whole, and for each individual variable:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   18.00   45.00   57.00   55.49   66.00   96.00       5
```

```
## [1] " "
```

```
##    1    2    3    4 NA's 
##  224  211   95  400   31
```

```
## [1] " "
```

```
##    1    2    3    4    5 NA's 
##  357   24  116  280  136   48
```

```
## [1] " "
```

```
##    1    2    3    4 NA's 
##   16   59  798   12   76
```

```
## [1] " "
```

```
##   0   1 
## 516 445
```

We also create a pairwise scatterplot to observe the relationships between individual variables (Figure 9):

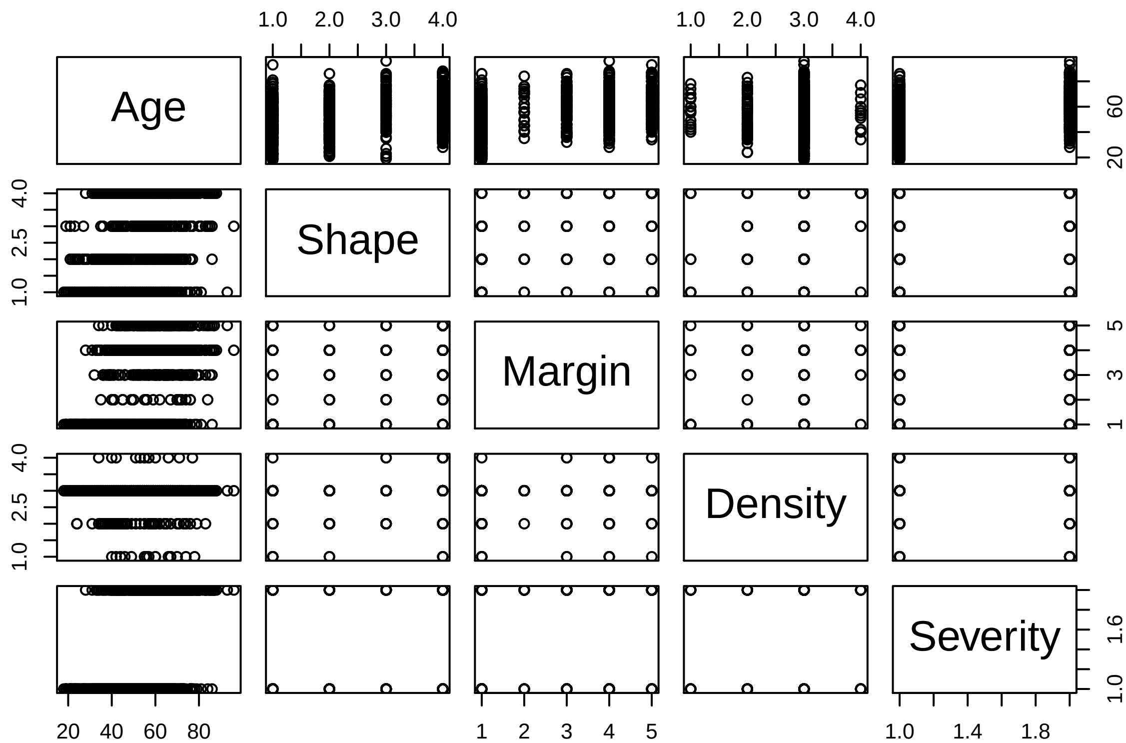

Figure 9: Pairwise scatterplot of Mammographic Mass Severity Data

There appears to be a weak, possibly linear, positive relationship between Age and Severity. There are no observable relationships between Severity and the other predictors.

Model Fitting and Model Selection

We now fit a logistic linear model (full.glm) to the data, with Severity as the response variable, and Age, Shape, Margin and Density as the predictor variables with interaction terms up to second order. To achieve a parsimony, iInsignificant terms (p-value>0.05) were removed by the backwards selection algorithmn. Our analysis showed that all the interactions terms and Density are not statistically significant at the 0.05. Details of each step of the model selection process are available in the Part B Appendix under "Model Fitting and Model Selection."

Justification of the final model

The following is the proposed final model.

```
##          term    estimate   std.error  statistic      p.value
## 1 (Intercept) -4.71954438 0.465771163 -10.132753 3.953566e-24
## 2         Age  0.05387869 0.007498812   7.184963 6.722520e-13
## 3      Shape2 -0.44784427 0.306327242  -1.461980 1.437467e-01
## 4      Shape3  0.49925125 0.364446283   1.369890 1.707213e-01
## 5      Shape4  1.24283749 0.324255926   3.832891 1.266463e-04
## 6     Margin2  1.58294301 0.539613781   2.933474 3.351917e-03
## 7     Margin3  1.26307280 0.342530673   3.687474 2.264916e-04
## 8     Margin4  1.54322611 0.294044551   5.248273 1.535316e-07
## 9     Margin5  2.03210483 0.362892142   5.599749 2.146627e-08
```

This model was obtained by first starting with a saturated model with all the two way interaction terms. Statistically insiginificant terms were removed via the backwards selection algorithmn at the 0.05 significance level. Hence the final model is the most parsimonious model with all the statistically significant predictors.

Interpretation of Parameters

Intercept: A woman of Age=0 with a mammogram of Shape=1 (round), Margin=1 (circumscribed) has log-odds=-4.7195 of having a malignant tumour.

Age: Holding all other variables constant, a one year increase in age increases the log-odds of having a malignant tumour by 0.05388.

Shape: Holding all other variables constant, having a lesion of Shape=2 (oval) increases the log-odds of having a malignant tumour by -.4478 compared to Shape=1 (round). For Shape=2 (lobular), th increase is 0.4992 and for Shape=3 (irregular), the increase is 1.2428, all relative to Shape=1.

Margin: Holding all other variables constant, having a lesion of Margin=2 (microlobulated) increases the log-odds of a malignant tumour by 1.5829 comapred to Margin=1 (circumscribed). For Margin =3 (obscured) the change in log-odds is 1.2631, Margin=4 (ill-defined) 1.5432 and Margin=5 (spiculated) 2.0321, all relative to Margin=1.

Predicting Probabilities and Interpretation

In this section, we use our final model (final.glm) to predict the probability of a specific patient, that is, a patient with given values for each of the predictor varaibles. Given that the response variable is defined to be 0 for benign (not cancerous) and 1 for malignant (cancerous), the fitted values lie between 0 and 1 and hence predict the probability for a given patient to have a malignant tumor.

We first fit the probabilties of each datapoint in the dataset based on the final model:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01736 0.12041 0.51500 0.47238 0.80361 0.96242
```

We can produce a histogram to visualise the overall distribution of probabilities:

```r
hist(probabilities)
```

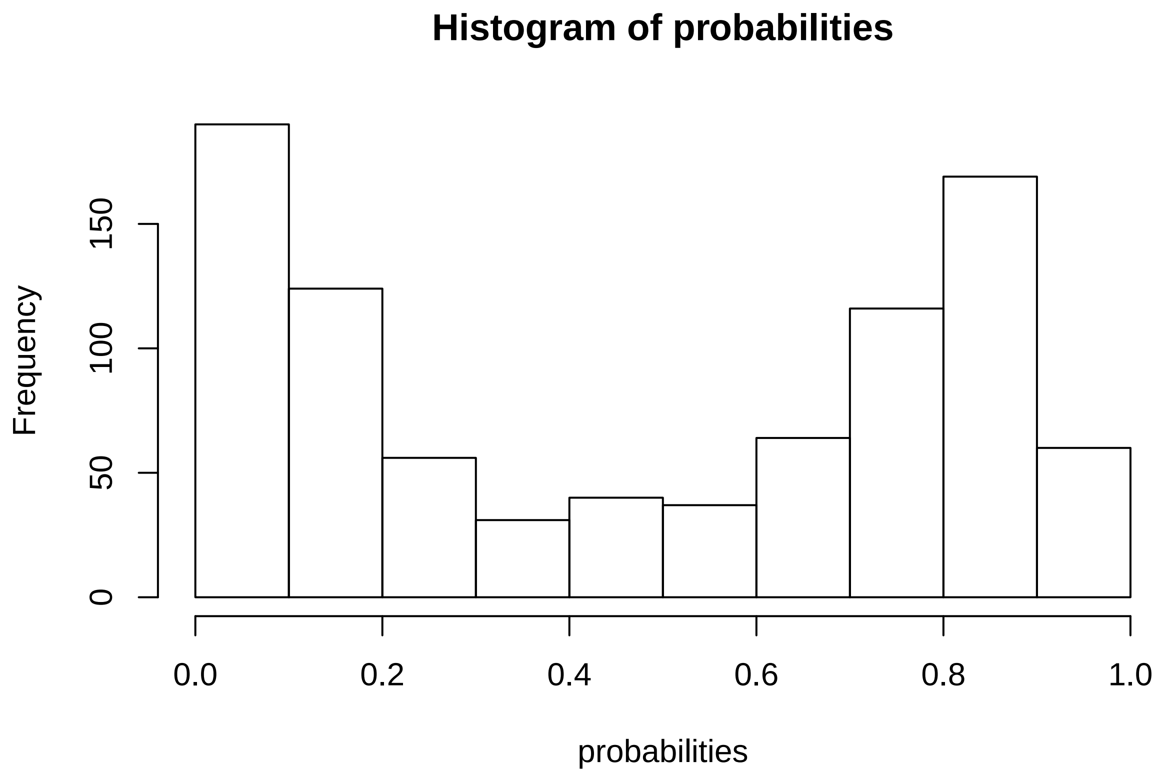

Figure 10: Histogram of Fitted Probabilities

Here we note that in general, it appears that most patients are either very likely, or very unlikely to have a malignant tumor. As a result, we might expect when predicting probabilities, that is most cases the predictions will be either very high or very low.

We can produce plots to visualise the probabilities for different levels of the predictor variables.

We first define a modified version of the mamo data, including only Age, Shape, Margin and Severity, and ignoring the missing values in order to be able to create valid plots:

We can now create plots of probabilities against Age, Shape and Margin:

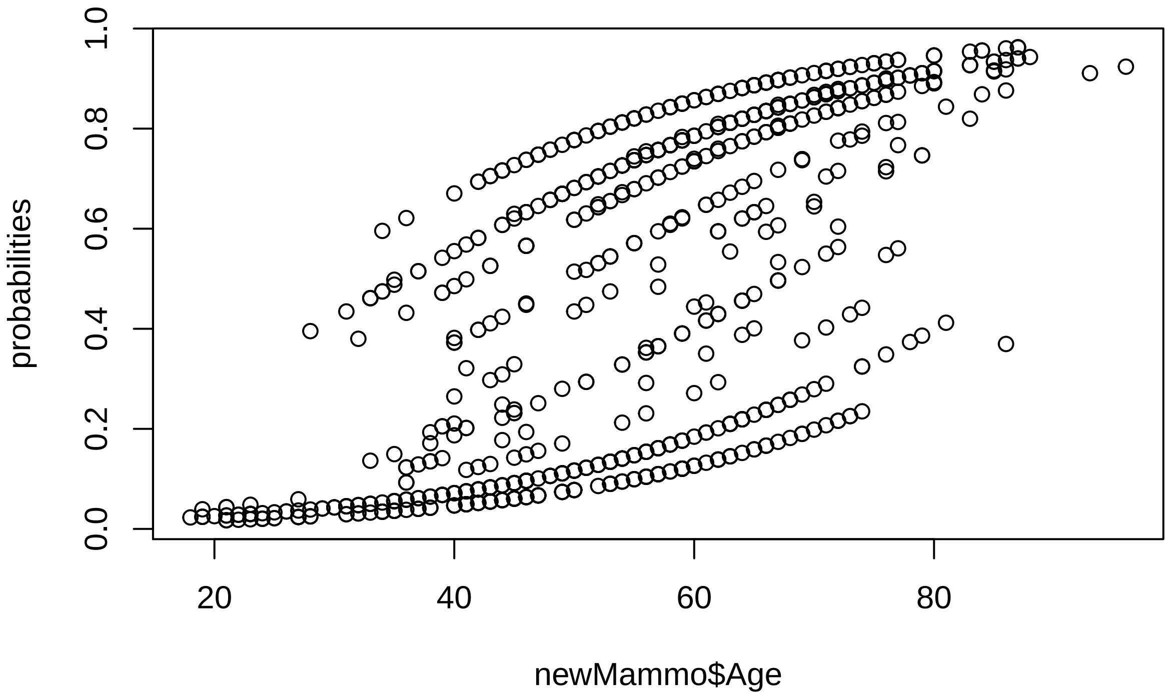

Figure 11: Probabilities against Age

Here we see that there appears to be a weak, positive relationship between age and the probability of having a malignant tumor, and it is difficult to say whether the relationship is linear or not.

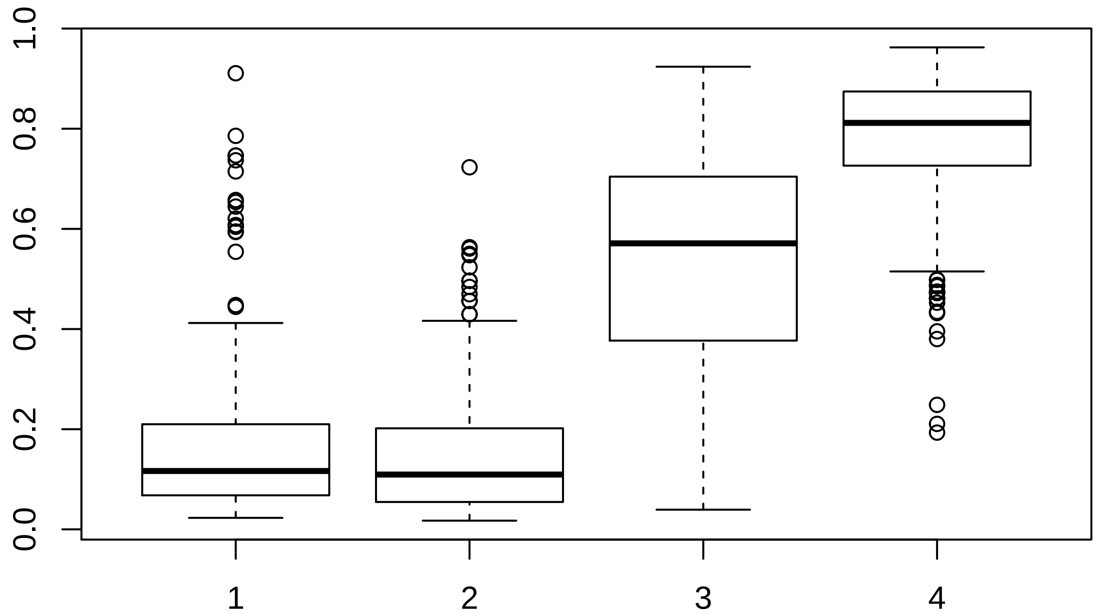

Figure 12: Probabilities against Shape

Here we see that as shape tends from the round, regular shape to a more irregular one, the predicted probabilities appear to increase in general.

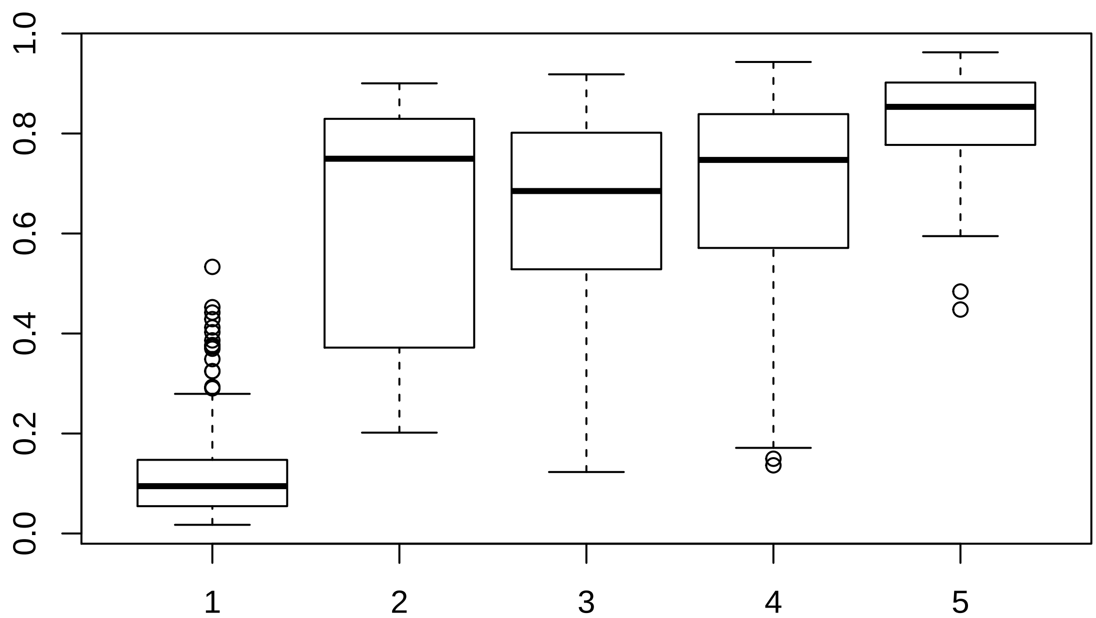

Figure 13: Probabilities against Margin

Here we see that as the margin tends from being well-defined to ill-defined, in general, the probability of the tumor being malignant seems to increase.

Having observed these relationships, we can now predict the probability of the tumor being malignant for a few specific patients. We do so for a patient at an age of 40, with Shape = 1 (round) and Margin = 1 (circumscribed). This is a patient which we would expect to have a relatively low probability, as they are quite young, and their tumor is quite regular in shape and margin.

The predicted probability is given by:

```
##          1
## 0.07146523
```

This probability of 0.07147 alligns well with what we would expect. We can interpret this to mean that from a large group of patients, those with an age of 40, round tumors and circumscribed margins, approximately 7% would have malignant tumors.

On the other end of the spectrum, we can predict the probability for a patient with an age of 80, Shape = 4 (irregular) and Margin = 5 (speculated), that is an older patient with an irregular and very much ill-defined tumor.

The predicted probability is given by:

```
##         1
## 0.9461242
```

This probability of 0.9461 also alligns well with what we would expect. This means, that for a large group of patients, we would expect that for patients of age 80, with irregularly shaped tumors and a spiculated margins, that approximately 95% would have malignant tumors.

To find what we would expect to be a more intermediate probability, we can then predict the probability for a patient with an age of 60, Shape = 2 (oval), Margin = 3 (obscured):

```
##         1
## 0.3381399
```

This probability of 0.3381 also makes sense intuitively, as the patient's age and margin values were much more intermediate, and the shape of their tumor is closer to regular end of the spectrum than the irregular end. This means, that in a large group of patients, we would expect that for patients of age 60, with oval shaped tumors and obscured margins, that approximately 34% would have malignant tumors.

Appendix

Part A

1.

Reading data into R:

```
child<-read.table("Child_Height.txt",header=T)
```

Creating pairwise scatterplot:

```
pairs(~Height+Weight+Length,data=child)
```

2.

Fitting linear models:

```
lm1<-lm(Length~Height+Weight, data=child)
lm2<-lm(Length~Height, data=child)
lm3<-lm(Length~Weight, data=child)
```

4.

Residuals plots for lm1:

```
par(mfrow=c(2,2))
plot(lm1)
```

Individual residuals vs. predictors plots for lm1:

```
par(mfrow=c(1,2))
res1<-rstudent(lm1)
fit<-fitted(lm1)
plot(child$Height,res1,main="Residuals vs height",pch=20)
abline(0,0,col="red")
plot(child$Weight,res1,main="Residuals vs weight",pch=20)
abline(0,0,col="red")
```

Residuals plots for lm2:

```
par(mfrow=c(2,2))
plot(lm2)
```

Individual residuals vs. predictors plots for lm2:

```
par(mfrow=c(1,2))
res1<-rstudent(lm2)
fit<-fitted(lm2)
plot(child$Height,res1,main="Residuals vs height",pch=20)
abline(0,0,col="red")
plot(child$Weight,res1,main="Residuals vs weight",pch=20)
abline(0,0,col="red")
```

Residuals plots for lm3:

```
par(mfrow=c(2,2))
plot(lm3)
```

Individual residuals vs. predictors plots for lm3:

```
par(mfrow=c(1,2))
res1<-rstudent(lm3)
fit<-fitted(lm3)
plot(child$Height,res1,main="Residuals vs height",pch=20)
abline(0,0,col="red")
plot(child$Weight,res1,main="Residuals vs weight",pch=20)
abline(0,0,col="red")
```

5.

Summary for lm1:

```
summary(lm1)
```

Summary for lm2:

```
summary(lm2)
```

Summary for lm3:

```
summary(lm3)
```

6.

(a)

Model matrices for lm2 and lm3:

```
lm2 <- lm(Length ~ Height, data = child)
lm3 <- lm(Length ~ Weight, data = child)
M2 <- model.matrix(lm2)
M3 <-  model.matrix(lm3)
M2
M3
```

(b)

Defining vector of ones, $\mathbf{x}_1$ and $\mathbf{x}_2$ for the Gram Schmidt Process, as well as the function for the norm of a vector:

```
one <- c(1,1,1,1,1,1,1,1,1,1,1,1) # intercept vector
x1 <- M2[,2] # vector of height values
x2 <- M3[,2] # vector of weight values
norm_vec <- function(x) sqrt(as.numeric(t(x) %*% x))
```

Finding an orthonormal basis for $\mathcal{L}_1$:

```r
v1 <- one / norm_vec(one)
v2_ <- x1 - as.numeric((t(x1) %*% v1)) * v1
v2 <- v2_ / norm_vec(v2_)
v2
```

Finding an orthonormal basis for $\mathcal{L}_2$:

```r
w1 <- one / norm_vec(one)
w2_ <- x2 - as.numeric((t(x2) %*% w1)) * w1
w2 <- w2_ / norm_vec(w2_)
w2
```

(c)

Computing the angle between the two spaces in **(b)**:

```r
inner <- t(v2) %*% (w2)
norm_v2 <- norm_vec(v2)
norm_w2 <- norm_vec(w2)
theta <- acos(inner/(norm_v2 * norm_w2))
theta
```