Name: Chee Kit Win

Student Number: 45589140

2. a) For a GLM, the distribution of V_i has to come from an exponential family. Since the normal distribution is part of the exponential family, once we have specified a suitable link function, we have a GLM. Here, we can use the log link. This link function gives. (From lecture notes)

$$\log(\mu_i) = \log(\beta_0 H_i^{\beta_1} D_i^{\beta_2})$$
$$= \log(\beta_0) + \beta_1 \log(H_i) + \beta_2 \log(D_i)$$

and this is the form we want since we have

$$g(\mu_i) = X_i^T \beta \quad \text{where} \quad g(\mu_i) = \log(\mu_i),$$
$$X_i^T = (1, \log H_i, \log D_i)$$
$$\beta = \begin{pmatrix} \log \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

b) $$\log(\mu_i) = -6.537 + 1.088 \log(H_i) + 1.997 \log(D_i)$$
$$\Rightarrow \mu_i = e^{-6.537} (H_i^{1.088})(D_i^{1.997})$$

$$V_i \sim N(e^{-6.537} H_i^{1.088} D_i^{1.997}, \sigma^2)$$

STAT3500 Assignment 2

Chee Kitt Win

8/31/2021

```
insects = read.csv("C:\\Users\\Owner\\Desktop\\UQ Year 3 Sem 2 Courses\\STAT3500\\Assignment
 2\\Insects.csv")
treedims = read.csv("C:\\Users\\Owner\\Desktop\\UQ Year 3 Sem 2 Courses\\STAT3500\\Assignment
2\\treedims.csv")
cancer = read.csv("C:\\Users\\Owner\\Desktop\\UQ Year 3 Sem 2 Courses\\STAT3500\\Assignment 2
\\cancer.csv", stringsAsFactors = TRUE)
library(lattice)
library(ggplot2)
```

Basic exploratory data analysis

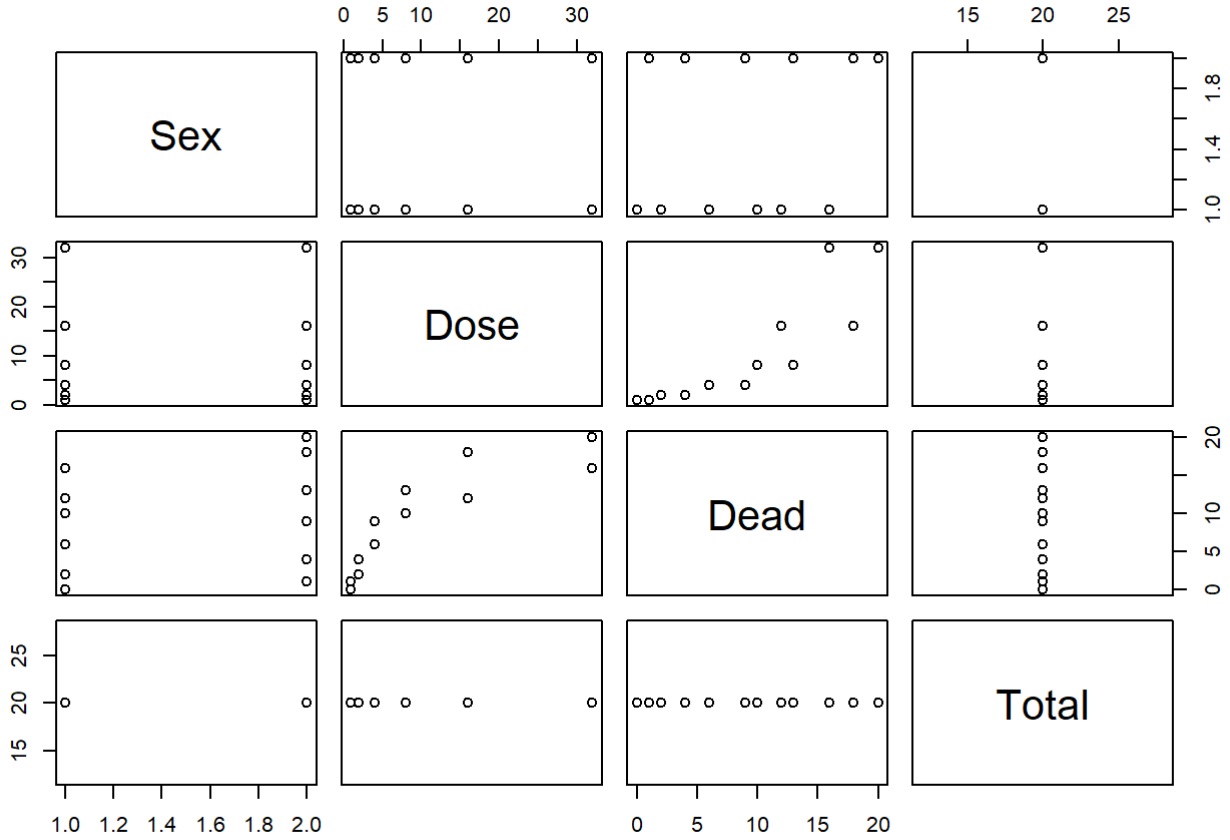
```
summary(insects)
```

```
##      Sex                 Dose           Dead           Total
##  Length:12          Min.   : 1.0   Min.   : 0.00   Min.   :20
##  Class :character   1st Qu.: 2.0   1st Qu.: 3.50   1st Qu.:20
##  Mode  :character   Median : 6.0   Median : 9.50   Median :20
##                     Mean   :10.5   Mean   : 9.25   Mean   :20
##                     3rd Qu.:16.0   3rd Qu.:13.75   3rd Qu.:20
##                     Max.   :32.0   Max.   :20.00   Max.   :20
```

```
head(insects)
```

```
##    Sex Dose Dead Total
## 1 male    1    1    20
## 2 male    2    4    20
## 3 male    4    9    20
## 4 male    8   13    20
## 5 male   16   18    20
## 6 male   32   20    20
```

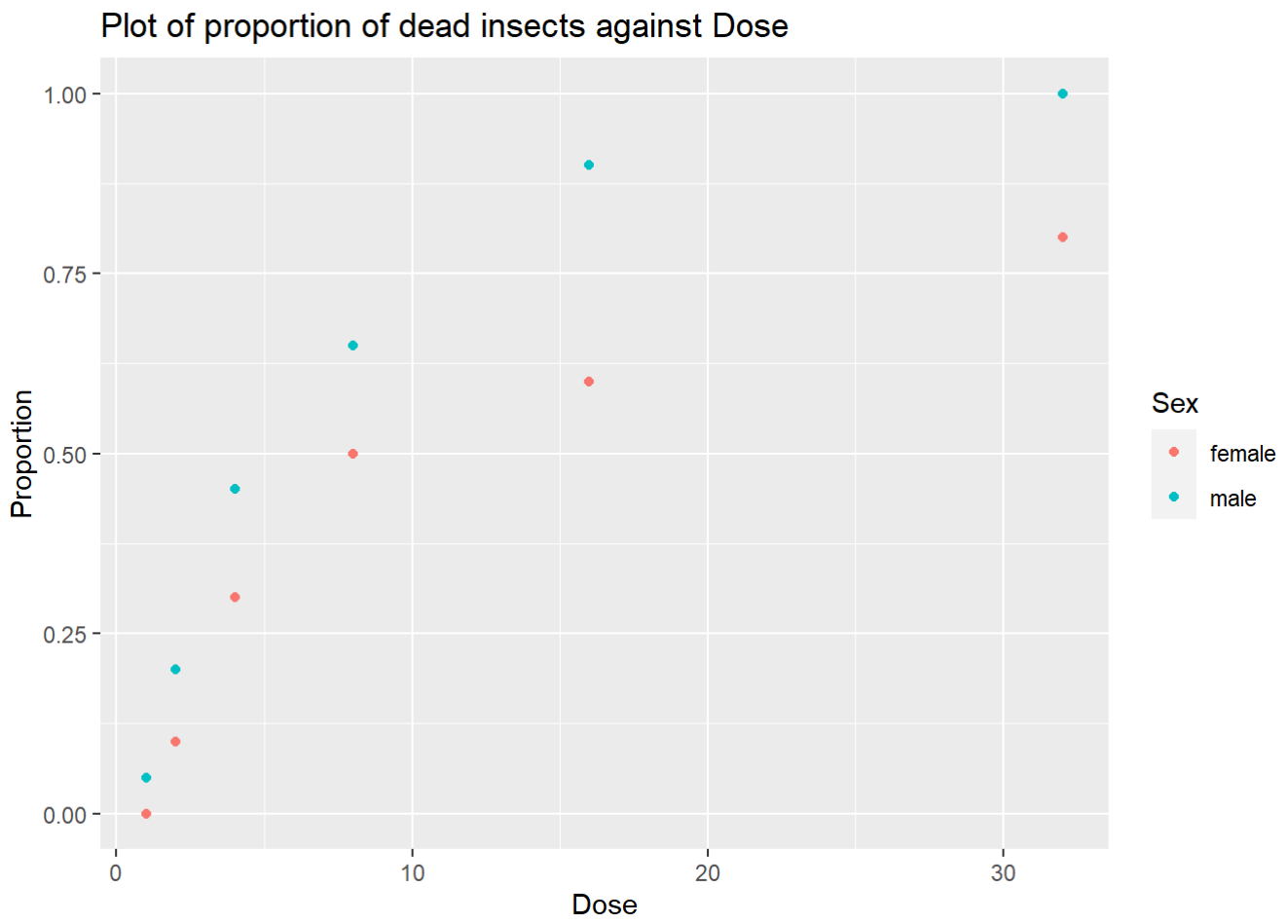
```
plot(insects)
```



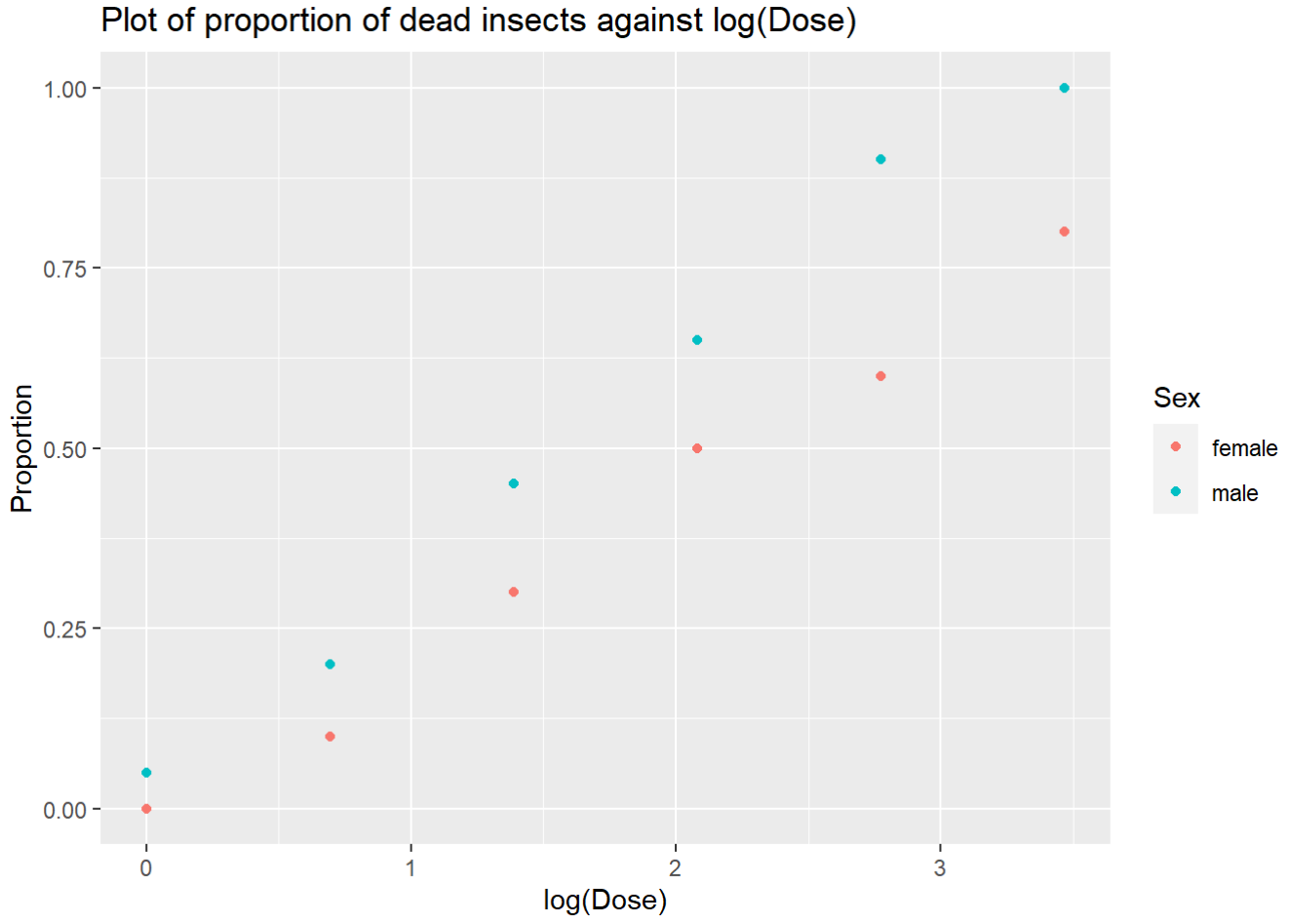
1. a)

For both sexes considered individually, the relationship between proportion of dead insects and dose looks very much like a log relationship. As expected, when we plot against logdose instead, we get a straight line and the relationship is roughly linear. From the plots, it seems like female insects might be more resistant to the chemical spray, but more analysis needs to be done to determine whether or not this is true.

```
insects$Proportion <- insects$Dead/insects$Total
qplot(Dose, Proportion, data = insects, color = Sex, main = "Plot of proportion of dead insec
ts against Dose")
```




```
qplot(log(Dose), Proportion, data = insects, color = Sex, main = "Plot of proportion of dead
 insects against log(Dose)")
```



```
# split_by_sex <- split(insects, insects$Sex)
# plot(split_by_sex$female$Dose,split_by_sex$female$Proportion, xlab = "Dose", ylab = "Proportion of dead female insects", main = "Plot of proportion of dead female insects vs dose")
# plot(split_by_sex$male$Dose,split_by_sex$male$Proportion, xlab = "Dose", ylab = "Proportion of dead male insects", main = "Plot of proportion of dead male insects vs dose")
# plot(log(split_by_sex$female$Dose),split_by_sex$female$Proportion, xlab = "log(Dose)", ylab = "Proportion of dead female insects", main = "Plot of proportion of dead female insects vs log(dose)")
# plot(log(split_by_sex$male$Dose),split_by_sex$male$Proportion, xlab = "log(Dose)", ylab = "Proportion of dead male insects",main = "Plot of proportion of dead male insects vs log(dose)")
```

1. b)

Below, we can see that the coefficient of the interaction term, sex*logdose has a p value of 0.191. This means that there is insufficient evidence to say that effect of logdose on the number of insects dead significantly differs between male and female insects. Despite the presence of the interaction term in the model, the coefficients of the other explanatory variables have very small p values which implies they have a significant effect on the number of insects alive (But this should checked again after dropping the interaction term). From the analysis of deviance table we see that log(Dose) has the largest effect on the response since the residual deviance drops by almost 95% after including log(Dose). Adding the interaction term into the model made hardly any reduction in the residual deviance, which indicates almost no improvement to the model. Applying a goodness of fit test, we get a p value of 1.753042e-13, which indicates a good fit.

```
insects$Alive = insects$Total - insects$Dead
response = cbind(insects$Dead, insects$Alive)
insects.int <- glm(response ~ 0 + Sex + log(Dose) + Sex * log(Dose), data = insects, family = binomial)
summary(insects.int)
```

```
##
## Call:
## glm(formula = response ~ 0 + Sex + log(Dose) + Sex * log(Dose),
##     family = binomial, data = insects)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.39849  -0.32094  -0.07592   0.38220   1.10375
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## Sexfemale          -2.9935     0.5527  -5.416 6.09e-08 ***
## Sexmale            -2.8186     0.5480  -5.143 2.70e-07 ***
## log(Dose)           1.3071     0.2411   5.422 5.89e-08 ***
## Sexmale:log(Dose)   0.5091     0.3895   1.307    0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 126.2269  on 12  degrees of freedom
## Residual deviance:   4.9937  on  8  degrees of freedom
## AIC: 43.104
##
## Number of Fisher Scoring iterations: 4
```

```
anova(insects.int, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: response
##
## Terms added sequentially (first to last)
##
##
##               Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                             12    126.227
## Sex            2    7.428        10    118.799  0.02438 *
## log(Dose)      1  112.042         9      6.757  < 2e-16 ***
## Sex:log(Dose)  1    1.763         8      4.994  0.18421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1 - pchisq(insects.int$null.deviance-insects.int$deviance,insects.int$df.null,insects.int$df.
residual)
```

```
## [1] 1.753042e-13
```

1. c)

The p value of the Chi squared statistic is 0.1842. The interpretation is that the model with the interaction term does not result in a significant improvement to the fit.

```
insects.main <- glm(response ~ 0 + Sex + log(Dose), data = insects, family = binomial)
summary(insects.main)
```

```
##
## Call:
## glm(formula = response ~ 0 + Sex + log(Dose), family = binomial,
##     data = insects)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.10540  -0.65343  -0.02225   0.48471   1.42944
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## Sexfemale  -3.4732     0.4685  -7.413 1.23e-13 ***
## Sexmale    -2.3724     0.3855  -6.154 7.56e-10 ***
## log(Dose)   1.5353     0.1891   8.119 4.70e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 126.2269  on 12  degrees of freedom
## Residual deviance:   6.7571  on  9  degrees of freedom
## AIC: 42.867
##
## Number of Fisher Scoring iterations: 4
```

```
anova(insects.main, test ="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: response
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                         12    126.227
## Sex        2    7.428        10    118.799  0.02438 *
## log(Dose)  1  112.042         9      6.757  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(insects.main, insects.int, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: response ~ 0 + Sex + log(Dose)
## Model 2: response ~ 0 + Sex + log(Dose) + Sex * log(Dose)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         9     6.7571
## 2         8     4.9937  1   1.7633   0.1842
```

1. d)

We have established that there is unlikely to be any interactions between sex and logdose, and so we can safely look at our main effects model. The significant P values for sexmale and sexfemale implies that yes, sex has a significant effect on the probability of an insect dying.

1. e)

For this part, the question said to compare the p value in the analysis of deviance in part c), with the p value reported in the summary of model output for part b). In my audio recorded answer, I took this to mean the p value of the interaction term in part b).

2. a)

See pdf file

2. b)

Graphically exploring the data

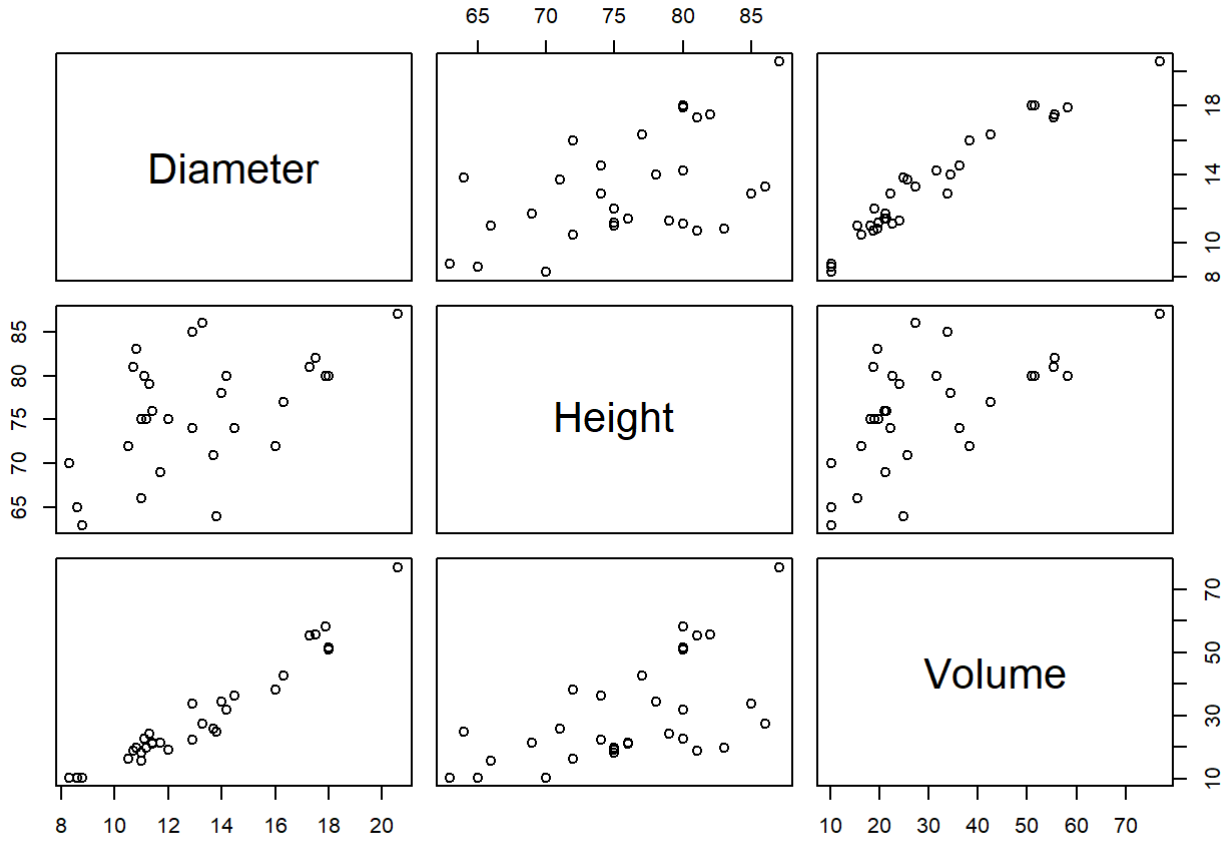
```
summary(treedims)
```

```
##     Diameter         Height       Volume
##  Min.   : 8.30   Min.   :63   Min.   :10.20
##  1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40
##  Median :12.90   Median :76   Median :24.20
##  Mean   :13.25   Mean   :76   Mean   :30.17
##  3rd Qu.:15.25   3rd Qu.:80   3rd Qu.:37.30
##  Max.   :20.60   Max.   :87   Max.   :77.00
```

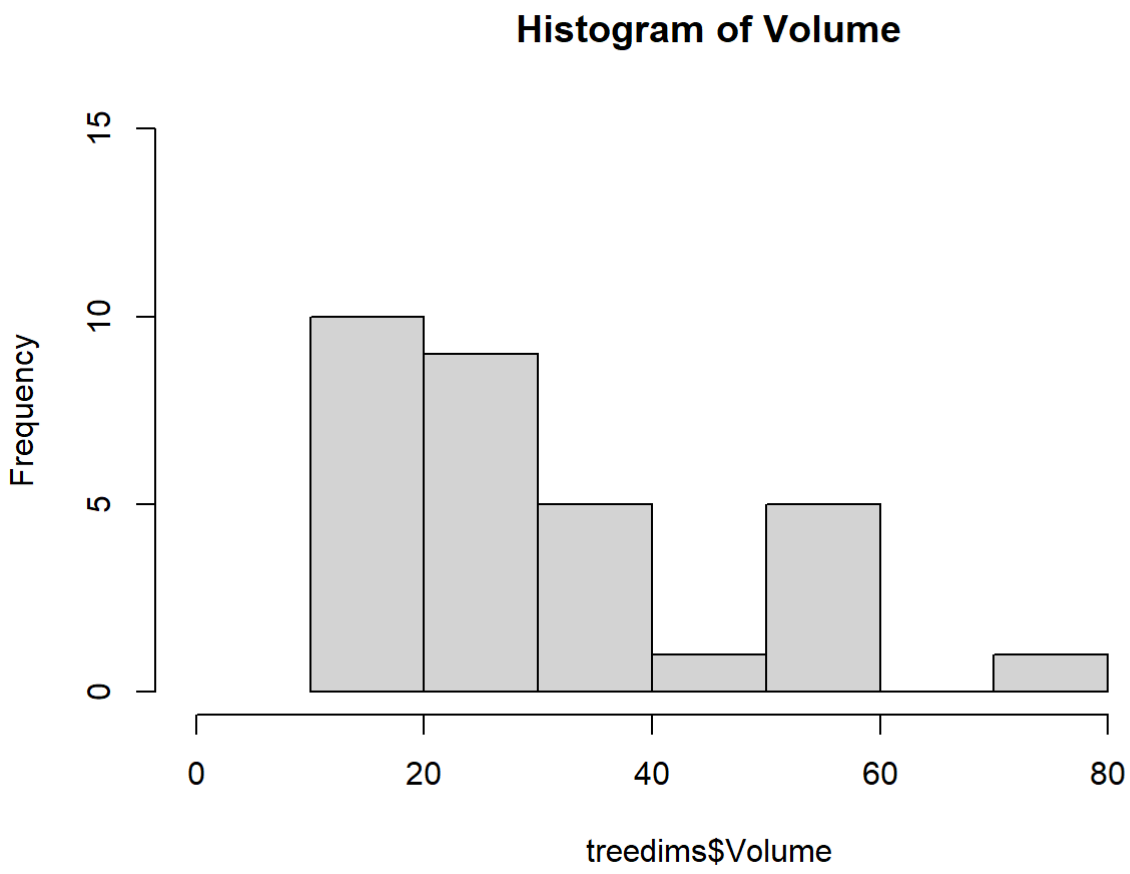
```
head(treedims)
```

```
##   Diameter Height Volume
## 1      8.3     70   10.3
## 2      8.6     65   10.3
## 3      8.8     63   10.2
## 4     10.5     72   16.4
## 5     10.7     81   18.8
## 6     10.8     83   19.7
```

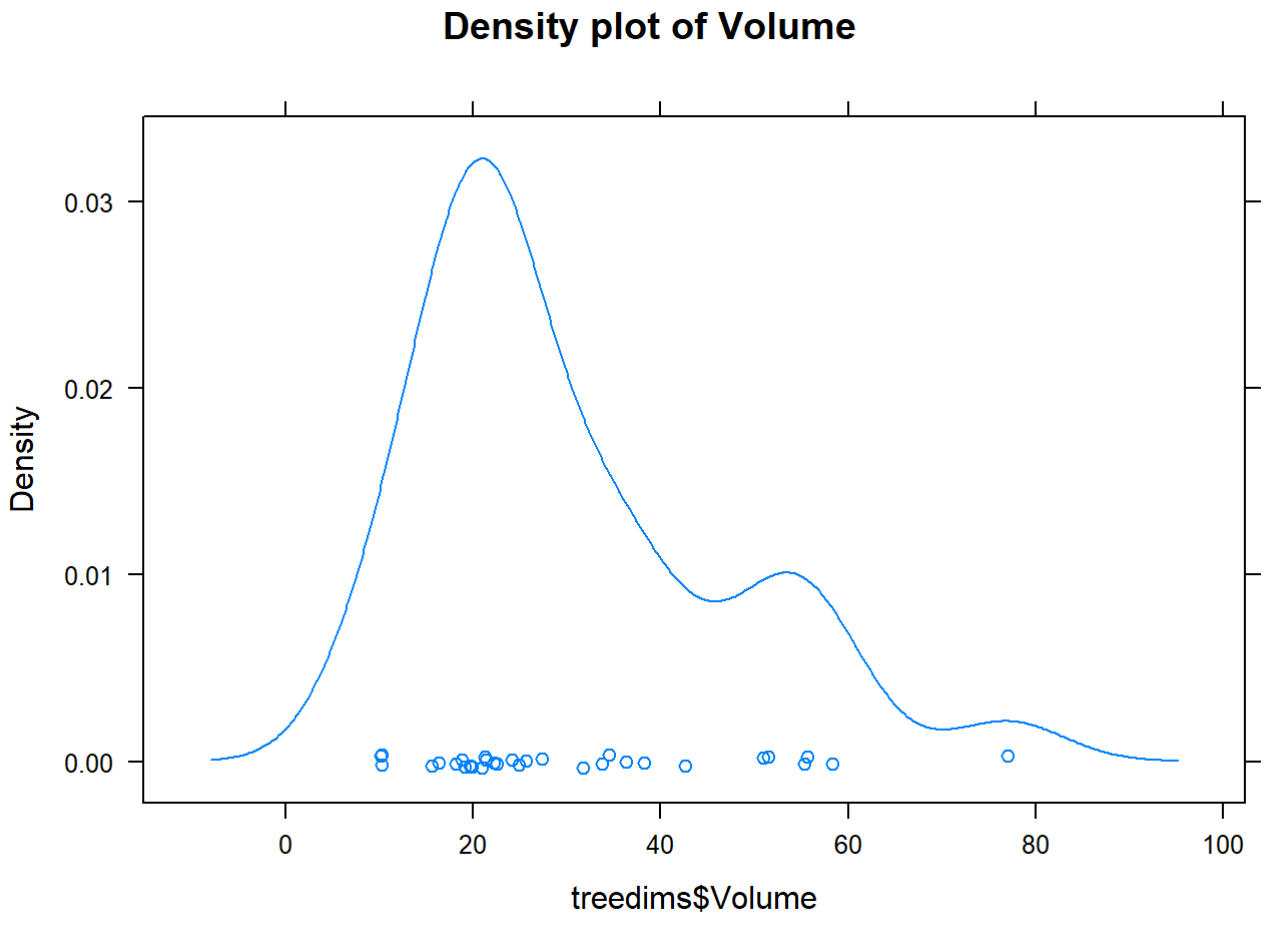
```
plot(treedims)
```




```
hist(treedims$Volume, main ="Histogram of Volume", xlim = c(0,90), ylim = c(0,15))
```




```
densityplot(treedims$Volume, main ="Density plot of Volume")
```



Fitting the model specified

```
treedims.main <- glm(Volume ~ log(Height) + log(Diameter), data = treedims, family = gaussian
(link =log))
summary(treedims.main)
```

```
##
## Call:
## glm(formula = Volume ~ log(Height) + log(Diameter), family = gaussian(link = log),
##     data = treedims)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.9080  -1.1817  -0.2101   1.7014   4.2551
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.53700    0.94352  -6.928 1.57e-07 ***
## log(Height)    1.08765    0.24216   4.491 0.000111 ***
## log(Diameter)  1.99692    0.08208  24.330  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.41642)
##
##     Null deviance: 8106.08  on 30  degrees of freedom
## Residual deviance:  179.66  on 28  degrees of freedom
## AIC: 150.44
##
## Number of Fisher Scoring iterations: 4
```

```
anova(treedims.main)
```

```
## Analysis of Deviance Table
##
## Model: gaussian, link: log
##
## Response: Volume
##
## Terms added sequentially (first to last)
##
##
##               Df Deviance Resid. Df Resid. Dev
## NULL                             30     8106.1
## log(Height)    1   2996.5        29     5109.6
## log(Diameter)  1   4930.0        28      179.7
```

Commenting on model output

The p values for log(height) and log(diameter) are both very significant implying that they after conducting a t-test, there is extremely strong evidence for rejecting the null (that the

coefficients = 0). log(Diameter) especially contributes to the fit of the model as can be seen by the large drop in residual deviance after log(diameter) is added to the deviance table.

For the expression for the fitted model, see the pdf file.

2. c)

We know the data is continuous and non-negative and from the histogram of Volume, we see that it is skewed to the right which suggests that a glm with a gamma distribution for V might be a good choice. We can use the log link, and similarly to what we did in 2 a), we will be able to obtain the form we want: g(mu) = X_transpose * B.

3.

First, some exploratory data analysis

```
summary(cancer)
```

```
##      counts         malignant        died     center
##  Min.   : 26.00   Min.   :0.0   Min.   :0.0   A:4
##  1st Qu.: 40.25   1st Qu.:0.0   1st Qu.:0.0   B:4
##  Median : 53.00   Median :0.5   Median :0.5
##  Mean   : 59.25   Mean   :0.5   Mean   :0.5
##  3rd Qu.: 76.25   3rd Qu.:1.0   3rd Qu.:1.0
##  Max.   :112.00   Max.   :1.0   Max.   :1.0
```

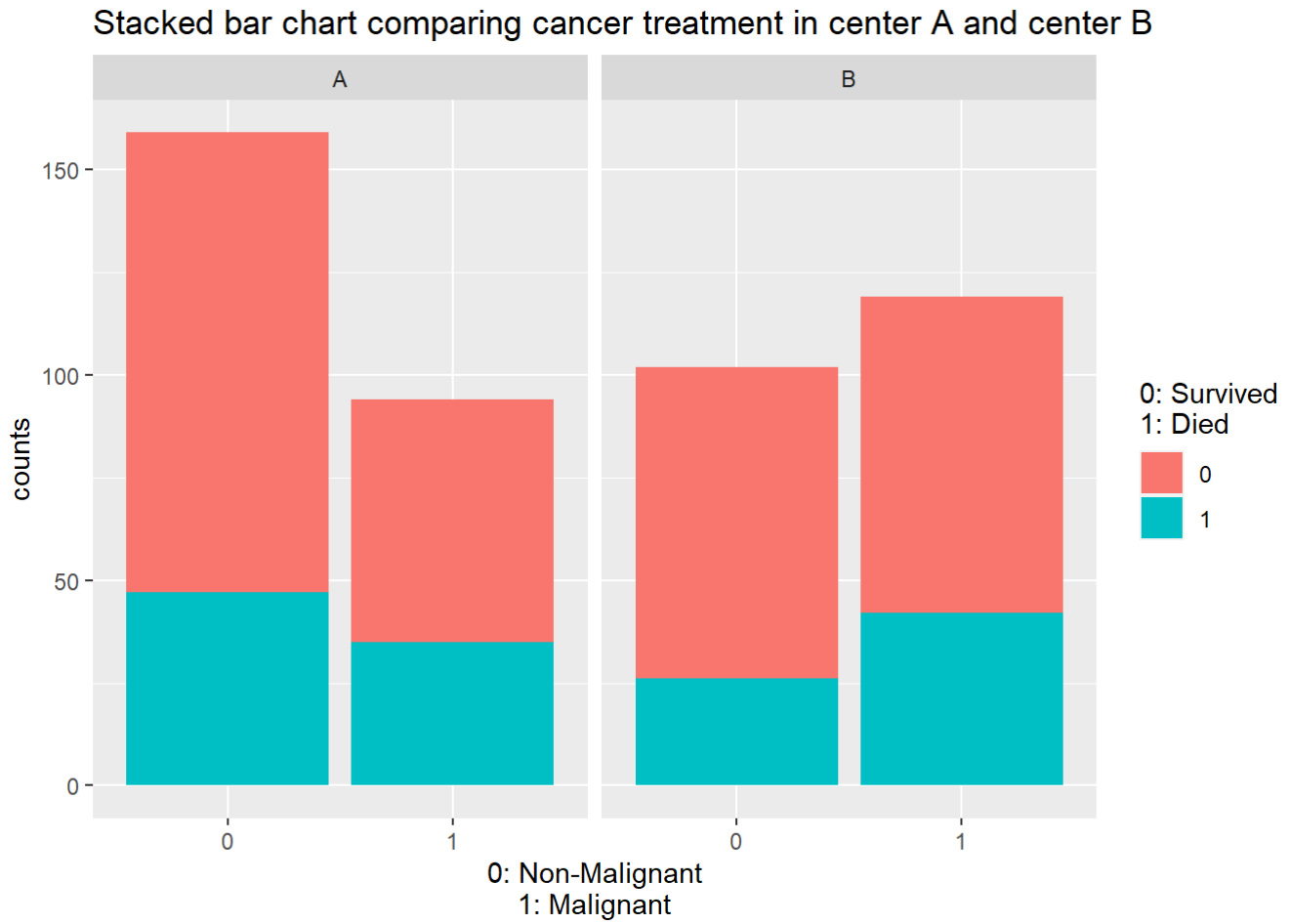
```
head(cancer)
```

```
##   counts malignant died center
## 1     35         1    1      A
## 2     42         1    1      B
## 3     59         1    0      A
## 4     77         1    0      B
## 5     47         0    1      A
## 6     26         0    1      B
```

```
library(ggplot2)
str(cancer)
```

```
## 'data.frame':    8 obs. of  4 variables:
##  $ counts   : int  35 42 59 77 47 26 112 76
##  $ malignant: int  1 1 1 1 0 0 0 0
##  $ died     : int  1 1 0 0 1 1 0 0
##  $ center   : Factor w/ 2 levels "A","B": 1 2 1 2 1 2 1 2
```

```
p <- ggplot(data = cancer, aes(x = factor(malignant),y = counts, fill = factor(died))) + face
t_wrap(~center) + geom_bar(stat="identity") + scale_color_discrete(labels = c(var1 = "Custom
 Value 1", var2 = "Custom Value 2", var3 = "Custom Value 3"))
p + xlab("0: Non-Malignant\n1: Malignant") + labs(title = "Stacked bar chart comparing cancer
treatment in center A and center B" ,fill = "0: Survived\n1: Died")
```



3. a)

Since we are fitting a saturated model, as expected the residual deviance is almost 0 with 0 degrees of freedom. The p value of malignant:centerB:died is 0.76645 which indicates that there is no evidence to suggest a third order interaction. We can then look at the second order terms. Only malignant:centerB has a significant P value, but it is possible that the other second order terms' p values will become significant after removing the third order interaction term (so we have to test this by dropping the third order interaction term).

```
glm.poisson = glm(counts ~ malignant*center*died, data = cancer, family = poisson)
summary(glm.poisson)
```

```
##
## Call:
## glm(formula = counts ~ malignant * center * died, family = poisson,
##     data = cancer)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            4.71850    0.09449  49.936  < 2e-16 ***
## malignant             -0.64096    0.16087  -3.984 6.76e-05 ***
## centerB               -0.38777    0.14862  -2.609  0.00908 **
## died                  -0.86835    0.17380  -4.996 5.84e-07 ***
## malignant:centerB      0.65403    0.22808   2.868  0.00414 **
## malignant:died         0.34616    0.27518   1.258  0.20842
## centerB:died          -0.20429    0.28605  -0.714  0.47513
## malignant:centerB:died  0.12034    0.40514   0.297  0.76645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 8.9970e+01  on 7  degrees of freedom
## Residual deviance: 1.8208e-14  on 0  degrees of freedom
## AIC: 62.602
##
## Number of Fisher Scoring iterations: 3
```

```
anova(glm.poisson, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: counts
##
## Terms added sequentially (first to last)
##
##
##                      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                     7     89.970
## malignant             1    4.869         6     85.101 0.0273417 *
## center                1    2.162         5     82.939 0.1414622
## died                  1   65.391         4     17.548 6.141e-16 ***
## malignant:center      1   13.328         3      4.220 0.0002614 ***
## malignant:died        1    3.619         2      0.600 0.0571074 .
## center:died           1    0.512         1      0.088 0.4744000
## malignant:center:died  1    0.088         0      0.000 0.7662720
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. b)

Model 2:

As mentioned in part a), the third order term is not significant as seen by its p value, so it is removed from the model. Refitting the model with the third order interaction term removed, we see that now only centerB:died is insignificant.

Model 3:

Removing centerB:died from the model, our variables are now all significant except for malignant:died, which is now borderline significant with a p value of 0.0571.

Model 4:

Refitting another model with malignant:died removed, we get a model with all variables having significant p values.

Since Model 3 has a lower AIC than Model 4, with a difference in AIC of 1.62, therefore I have chosen Model 3 as my final model.

```
glm.poisson2 = glm(counts ~ (malignant + center + died)^2, data = cancer, family = poisson)
summary(glm.poisson2)
```

```
##
## Call:
## glm(formula = counts ~ (malignant + center + died)^2, family = poisson,
##     data = cancer)
##
## Deviance Residuals:
##        1         2         3         4         5         6         7         8
## -0.12334   0.11405   0.09606  -0.08347   0.10774  -0.14277  -0.06928   0.08456
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        4.72504    0.09159  51.589  < 2e-16 ***
## malignant         -0.66003    0.14794  -4.462 8.14e-06 ***
## centerB           -0.40402    0.13839  -2.919 0.003507 **
## died              -0.89065    0.15747  -5.656 1.55e-08 ***
## malignant:centerB  0.69232    0.18847   3.673 0.000239 ***
## malignant:died     0.40177    0.20157   1.993 0.046236 *
## centerB:died      -0.14440    0.20212  -0.714 0.474957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 89.97022  on 7  degrees of freedom
## Residual deviance:  0.08836  on 1  degrees of freedom
## AIC: 60.691
##
## Number of Fisher Scoring iterations: 3
```

```
anova(glm.poisson2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: counts
##
## Terms added sequentially (first to last)
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                 7     89.970
## malignant         1    4.869         6     85.101 0.0273417 *
## center            1    2.162         5     82.939 0.1414622
## died              1   65.391         4     17.548 6.141e-16 ***
## malignant:center  1   13.328         3      4.220 0.0002614 ***
## malignant:died    1    3.619         2      0.600 0.0571074 .
## center:died       1    0.512         1      0.088 0.4744000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
glm.poisson3 = glm(counts ~ malignant + center + died + malignant*center + malignant*died, da
ta = cancer, family = poisson)
summary(glm.poisson3)
```

```
##
## Call:
## glm(formula = counts ~ malignant + center + died + malignant *
##     center + malignant * died, family = poisson, data = cancer)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##  0.1739  -0.1560  -0.1319   0.1166   0.3757  -0.4807  -0.2372   0.2933
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        4.74083    0.08819  53.758  < 2e-16 ***
## malignant         -0.64617    0.14517  -4.451 8.54e-06 ***
## centerB           -0.44393    0.12686  -3.499 0.000466 ***
## died              -0.94598    0.13790  -6.860 6.90e-12 ***
## malignant:centerB  0.67976    0.18744   3.626 0.000287 ***
## malignant:died     0.37713    0.19839   1.901 0.057303 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 89.97022  on 7  degrees of freedom
## Residual deviance:  0.60007  on 2  degrees of freedom
## AIC: 59.202
##
## Number of Fisher Scoring iterations: 3
```

```
anova(glm.poisson3, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: counts
##
## Terms added sequentially (first to last)
##
##
##                 Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                7     89.970
## malignant         1    4.869         6     85.101 0.0273417 *
## center            1    2.162         5     82.939 0.1414622
## died              1   65.391         4     17.548 6.141e-16 ***
## malignant:center  1   13.328         3      4.220 0.0002614 ***
## malignant:died    1    3.619         2      0.600 0.0571074 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1 - pchisq(glm.poisson3$null.deviance-glm.poisson3$deviance,glm.poisson3$df.null,glm.poisson3
$df.residual)
```

```
## [1] 1.114664e-13
```

```
glm.poisson4 = glm(counts ~ malignant + center + died + malignant*center, data = cancer, fami
ly = poisson)
summary(glm.poisson4)
```

```
##
## Call:
## glm(formula = counts ~ malignant + center + died + malignant *
##     center, family = poisson, data = cancer)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##  0.9367   0.6945  -0.6646  -0.4858  -0.4728  -1.1442   0.3165   0.7410
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        4.68844    0.08524  55.002  < 2e-16 ***
## malignant         -0.52561    0.13011  -4.040 5.35e-05 ***
## centerB           -0.44393    0.12686  -3.499 0.000466 ***
## died              -0.77011    0.09876  -7.798 6.29e-15 ***
## malignant:centerB  0.67976    0.18744   3.626 0.000287 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 89.9702  on 7  degrees of freedom
## Residual deviance:  4.2195  on 3  degrees of freedom
## AIC: 60.822
##
## Number of Fisher Scoring iterations: 4
```

```
anova(glm.poisson4, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: counts
##
## Terms added sequentially (first to last)
##
##
##                 Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                7     89.970
## malignant         1    4.869         6     85.101 0.0273417 *
## center            1    2.162         5     82.939 0.1414622
## died              1   65.391         4     17.548 6.141e-16 ***
## malignant:center  1   13.328         3      4.220 0.0002614 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
1 - pchisq(glm.poisson4$null.deviance-glm.poisson4$deviance,glm.poisson4$df.null,glm.poisson4
$df.residual)
```

```
## [1] 3.391731e-12
```

3. c)

The estimated dispersion parameter is 0.298. This means that Model 3 had severe underdispersion (due to the varriance being constrained by the mean, since with a poisson distribution, mean = variance). By relaxing this condition, the quasipoisson distribution has accounted for the different variance so that the model fits better.

```
glm.quasipoisson = glm(counts ~ malignant + center + died + malignant*center + malignant*die
d, data = cancer, family = quasipoisson)
summary(glm.quasipoisson)
```

```
##
## Call:
## glm(formula = counts ~ malignant + center + died + malignant *
##     center + malignant * died, family = quasipoisson, data = cancer)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##  0.1739  -0.1560  -0.1319   0.1166   0.3757  -0.4807  -0.2372   0.2933
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.74083    0.04816  98.443 0.000103 ***
## malignant         -0.64617    0.07927  -8.151 0.014720 *
## centerB           -0.44393    0.06928  -6.408 0.023497 *
## died              -0.94598    0.07531 -12.561 0.006278 **
## malignant:centerB  0.67976    0.10236   6.641 0.021932 *
## malignant:died     0.37713    0.10834   3.481 0.073535 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.2982123)
##
##     Null deviance: 89.97022  on 7  degrees of freedom
## Residual deviance:  0.60007  on 2  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
```

```
1 - pchisq(glm.quasipoisson$null.deviance-glm.quasipoisson$deviance,glm.quasipoisson$df.null,
glm.quasipoisson$df.residual)
```

```
## [1] 1.114664e-13
```