

HW 5

Ted Henson

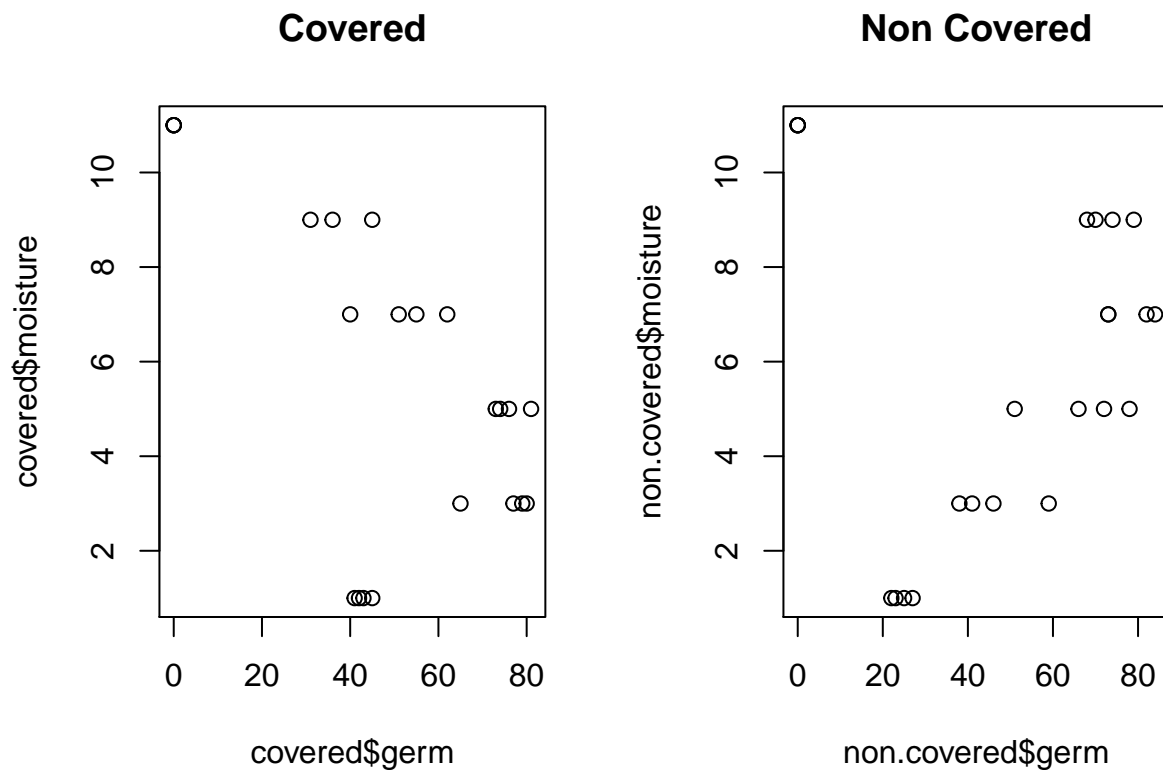
2/24/2020

Chapter 3 Problem 3, page 66

a)

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.2.1      v purrr  0.3.3
## v tibble  2.1.3      v dplyr  0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

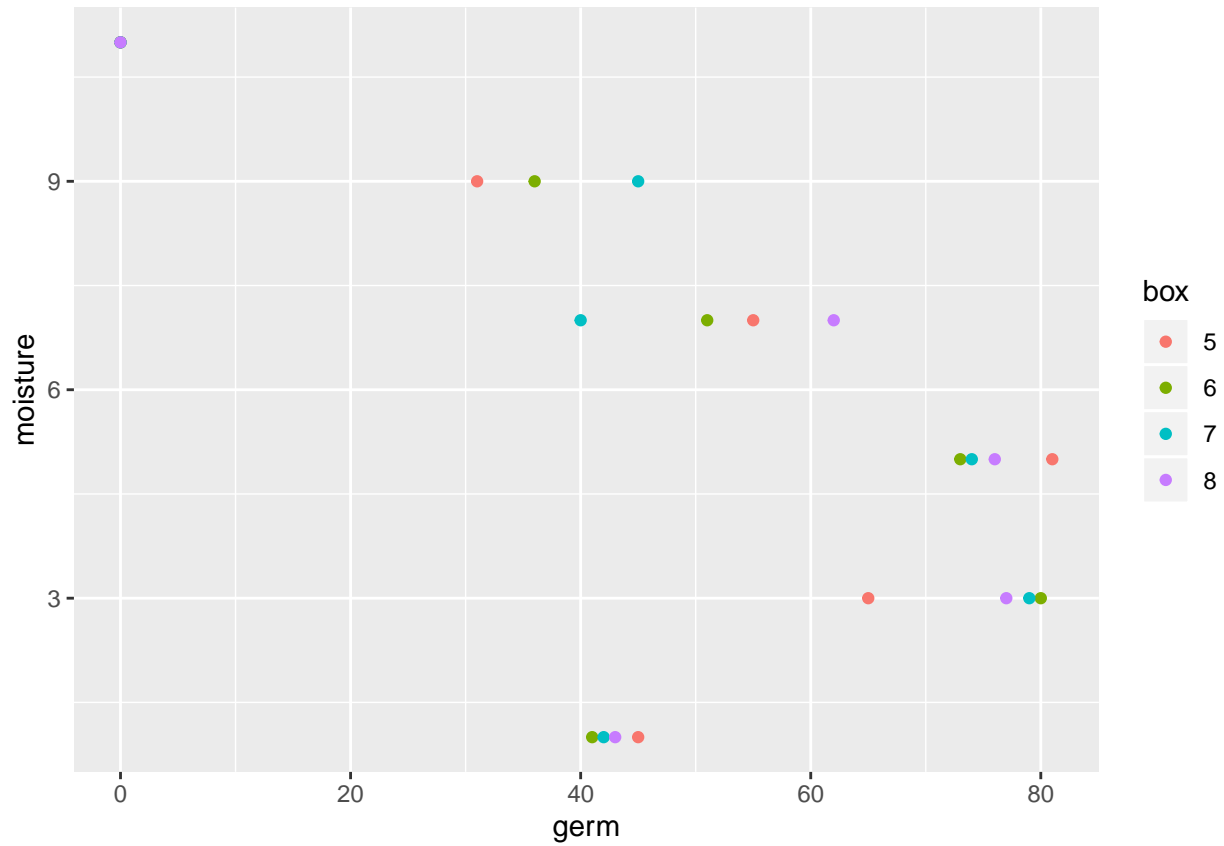
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

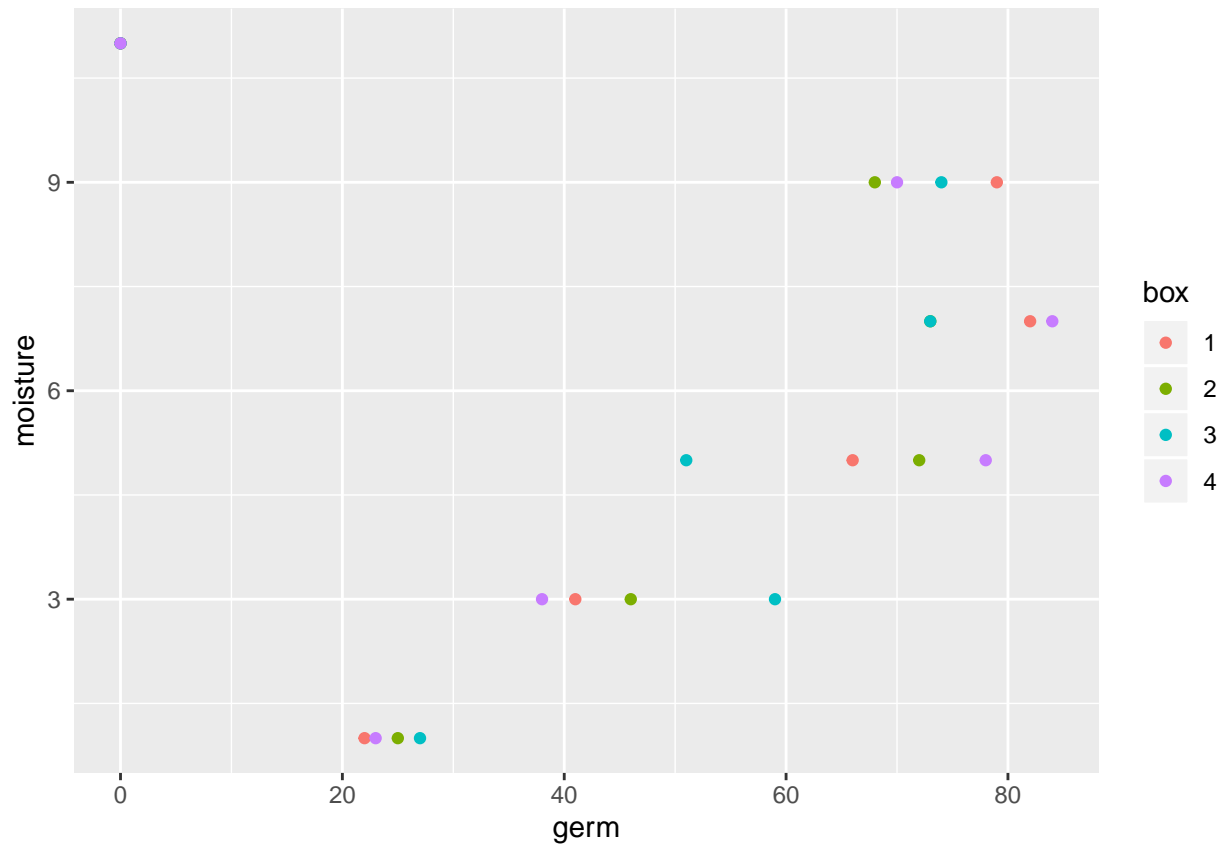


The moisture and germination percentage appear to be negatively correlated when the box is covered, but positively correlated when the box is uncovered.

b)

```
## Warning: Removed 1 rows containing missing values (geom_point).
```





Yes, it appears there may be a box effect. Seeds within the same box follow a linear trajectory depending on covered versus uncovered.

c)

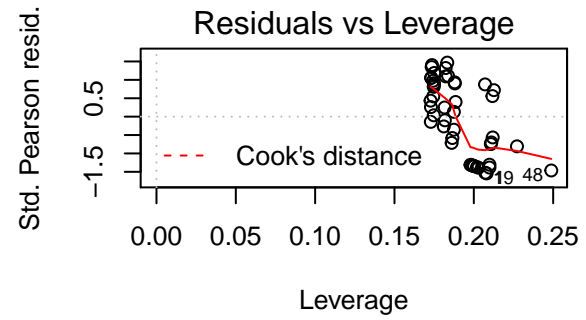
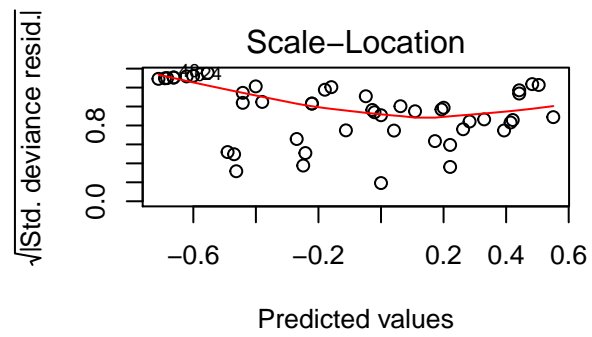
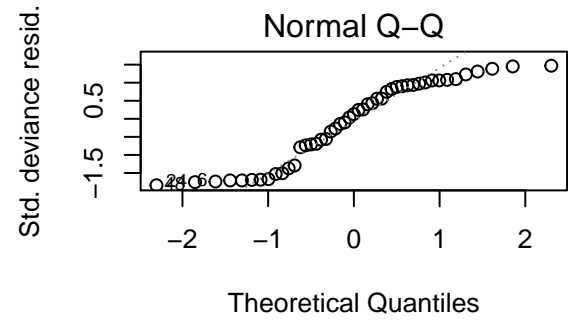
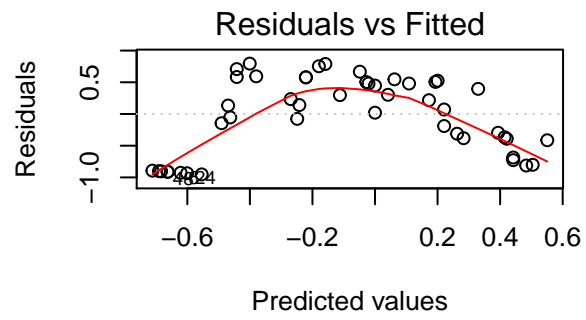
```
##
## Call:
## glm(formula = germ ~ box + moisture, family = quasibinomial,
##      data = seeds.new)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95285  -0.55046   0.07063   0.50465   0.79405
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.59387    0.58755   1.011  0.3185
## box2        -0.04149    0.70283  -0.059  0.9532
## box3        -0.04149    0.70283  -0.059  0.9532
## box4         0.02072    0.70244   0.030  0.9766
## box5        -0.09001    0.70335  -0.128  0.8988
## box6        -0.06227    0.70303  -0.089  0.9299
## box7        -0.06920    0.70310  -0.098  0.9221
## box8         0.06730    0.73750   0.091  0.9278
## moisture    -0.11049    0.05266  -2.098  0.0426 *
## ---
```

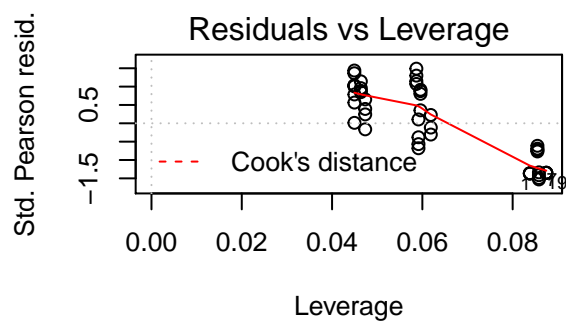
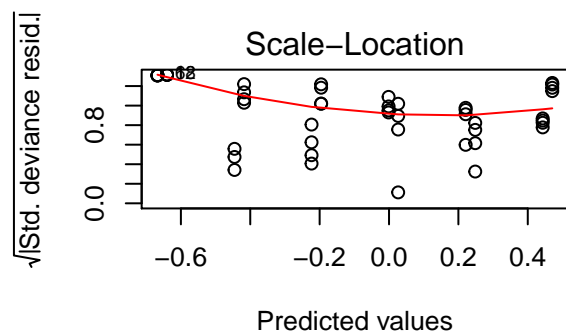
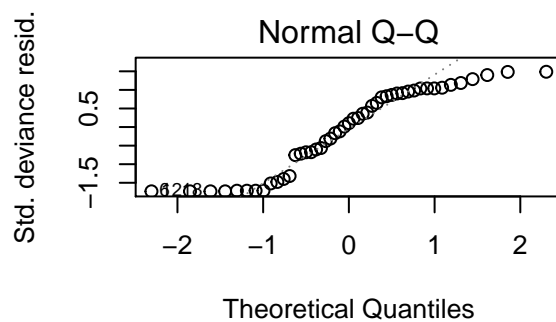
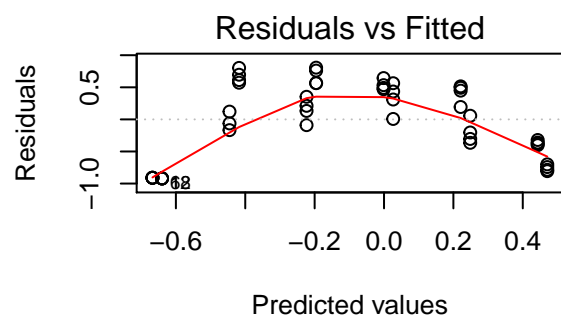
```

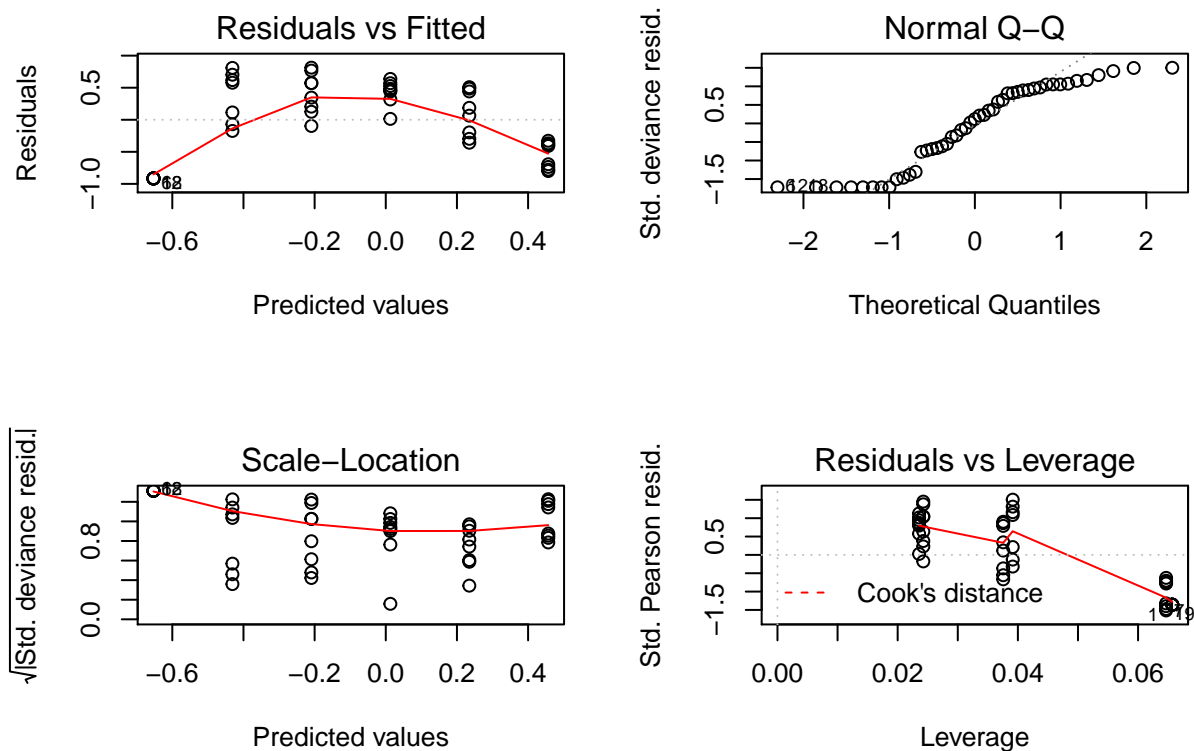
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.3571225)
##
##      Null deviance: 17.910  on 46  degrees of freedom
## Residual deviance: 16.244  on 38  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = germ ~ covered + moisture, family = quasibinomial,
##      data = seeds.new)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92035  -0.55103   0.05695   0.50362   0.80597
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.58170     0.37170   1.565  0.1248
## coveredyes  -0.02760     0.33058  -0.083  0.9338
## moisture    -0.11106     0.04894  -2.269  0.0282 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.3090345)
##
##      Null deviance: 17.910  on 46  degrees of freedom
## Residual deviance: 16.268  on 44  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = germ ~ moisture, family = quasibinomial, data = seeds.new)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91546  -0.55099   0.06376   0.49703   0.81244
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.56770     0.32786   1.732  0.0902 .
## moisture    -0.11097     0.04838  -2.294  0.0265 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.3022393)
##
##      Null deviance: 17.91  on 46  degrees of freedom
## Residual deviance: 16.27  on 45  degrees of freedom
## AIC: NA

```

```
##
## Number of Fisher Scoring iterations: 3
```







The model with just moisture should be chosen since the plots shown above all look about the same, so the model with the least number of predictors should be chosen in this scenario. All predictors aside from the moisture had large p values.

d)

```
## Analysis of Deviance Table
##
## Model 1: germ ~ moisture
## Model 2: germ ~ box + moisture
##   Resid. Df Resid. Dev Df Deviance
## 1      45      16.270
## 2      38      16.244  7  0.025703
## [1] 1
```

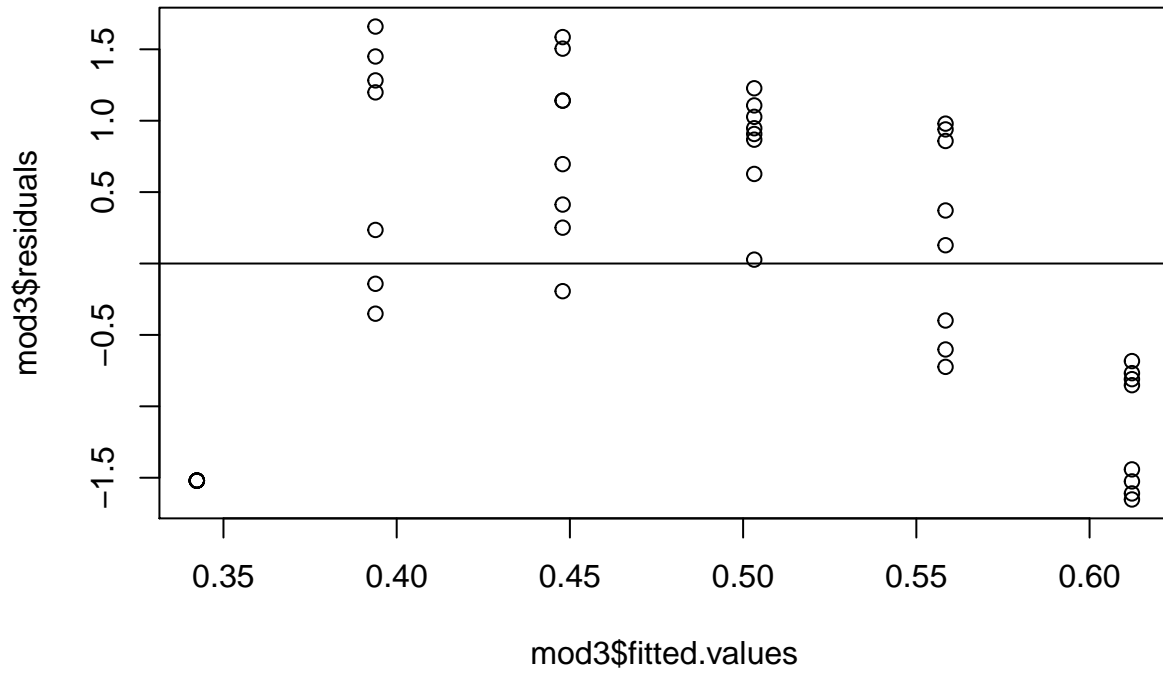
The deviance does not decrease by much with the inclusion of the box parameter. The chisq test confirms with a p value of 1 that it is extremely unlikely that the box parameter is statistically significant.

e)

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

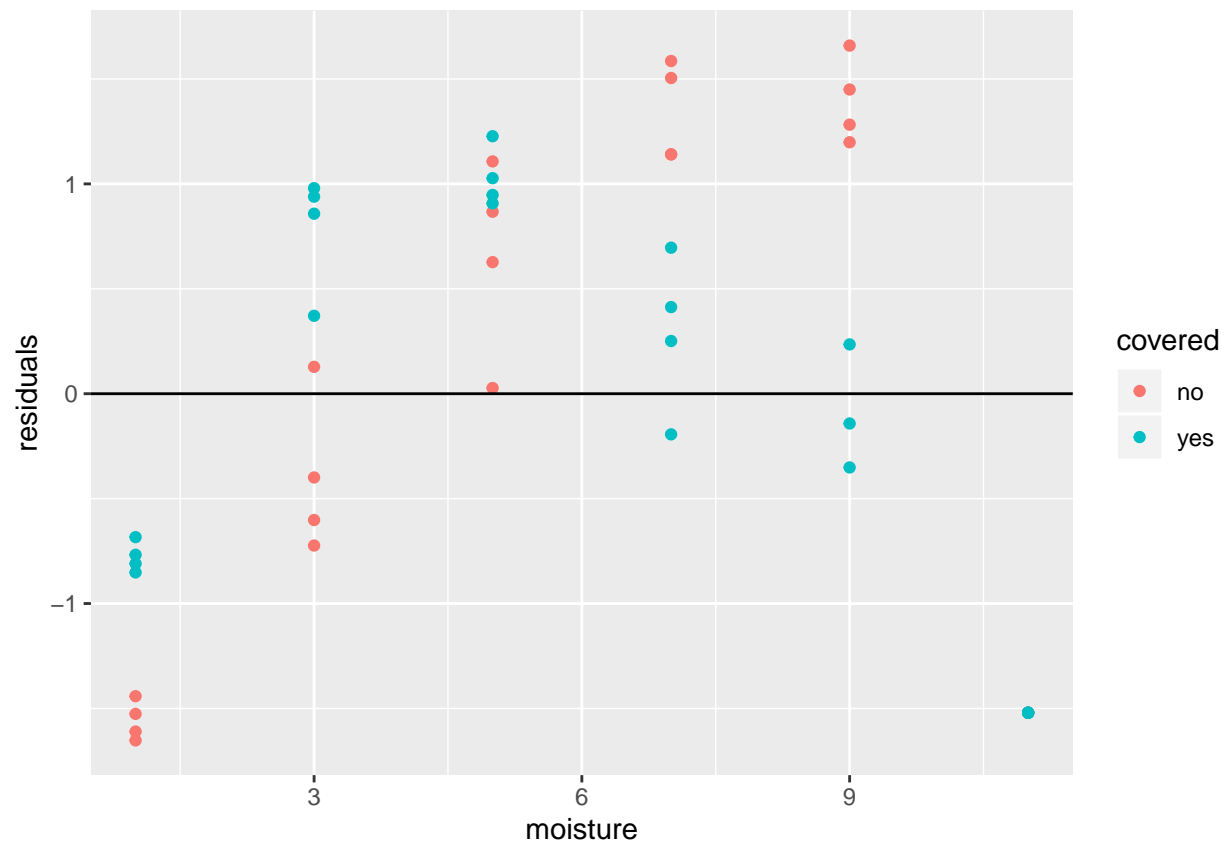
The maximum germination occurs at a moisture of 1 for both covered and uncovered boxes.

f)



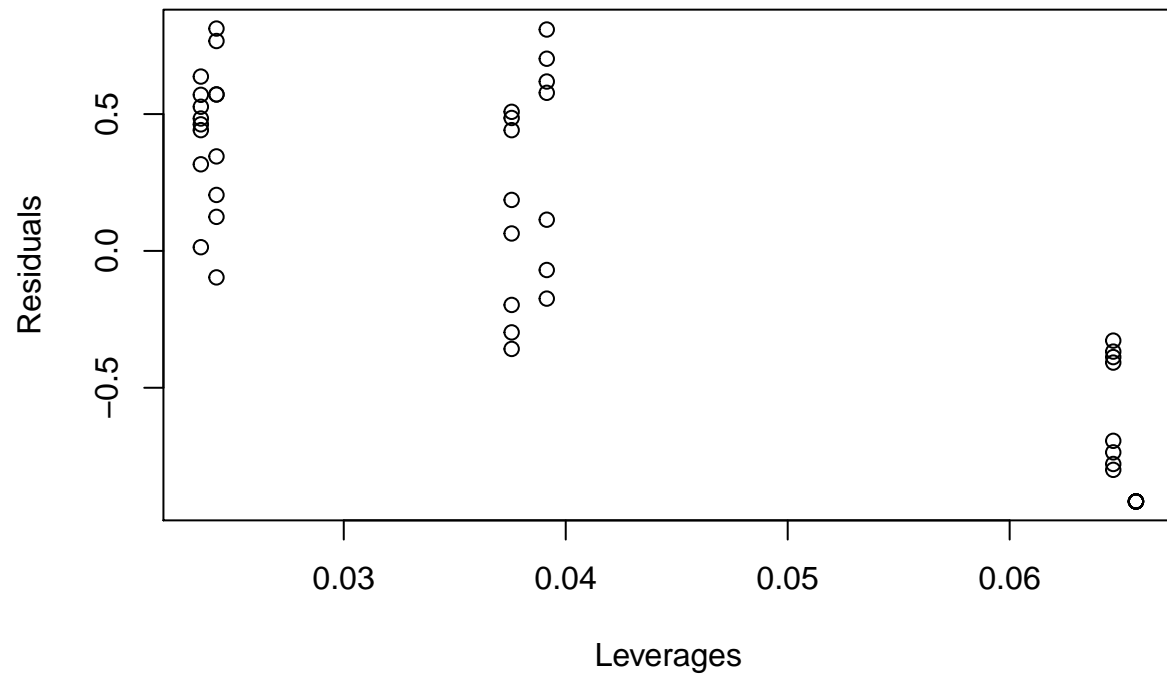
The residuals are strictly negative for values fitted at the maximum germination value. In the middle quartiles, the residuals are mostly positive, but some negative. There is one particularly large negative residual for the smallest fitted value.

g)



As with the previous plot, the residuals appear to be larger at extreme values of moisture. This plot shows that residuals may be larger for uncovered boxes, but there is not substantial visual evidence or statistical evidence based on this plot.

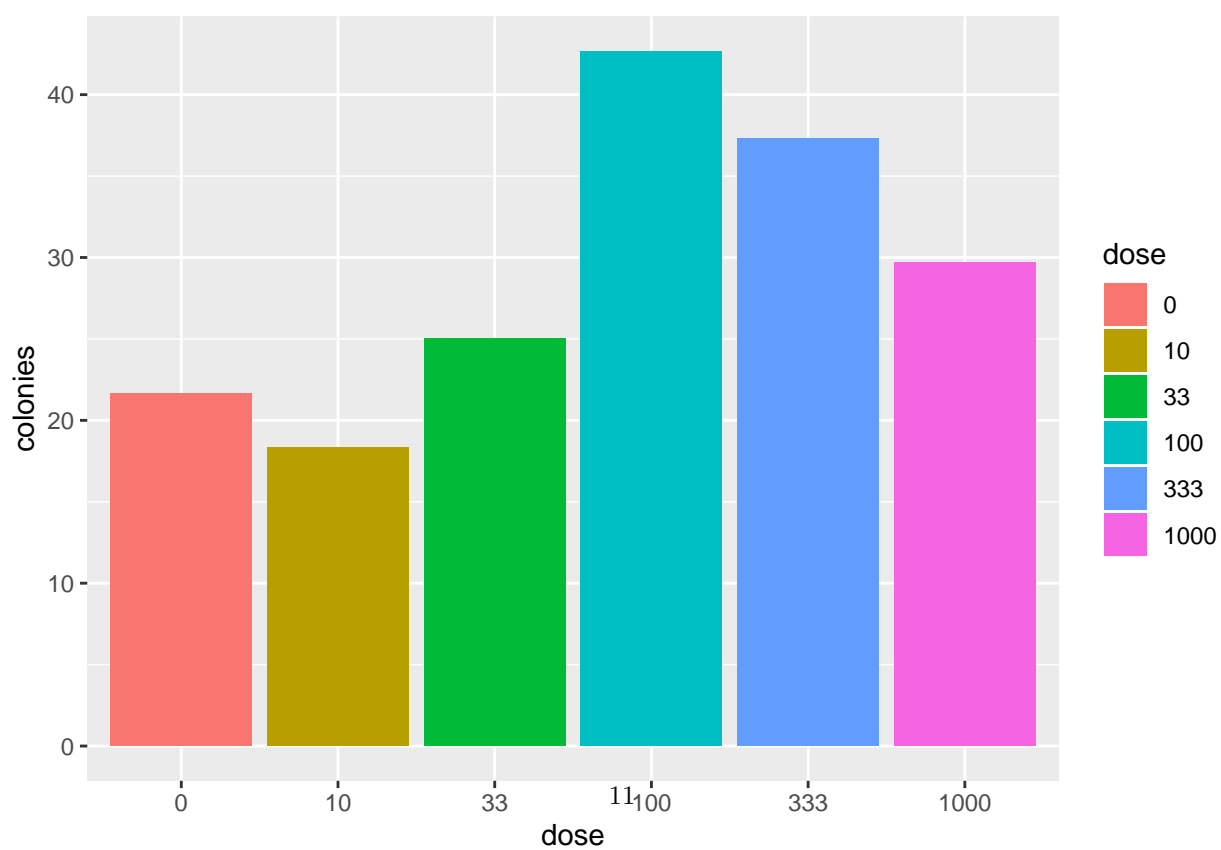
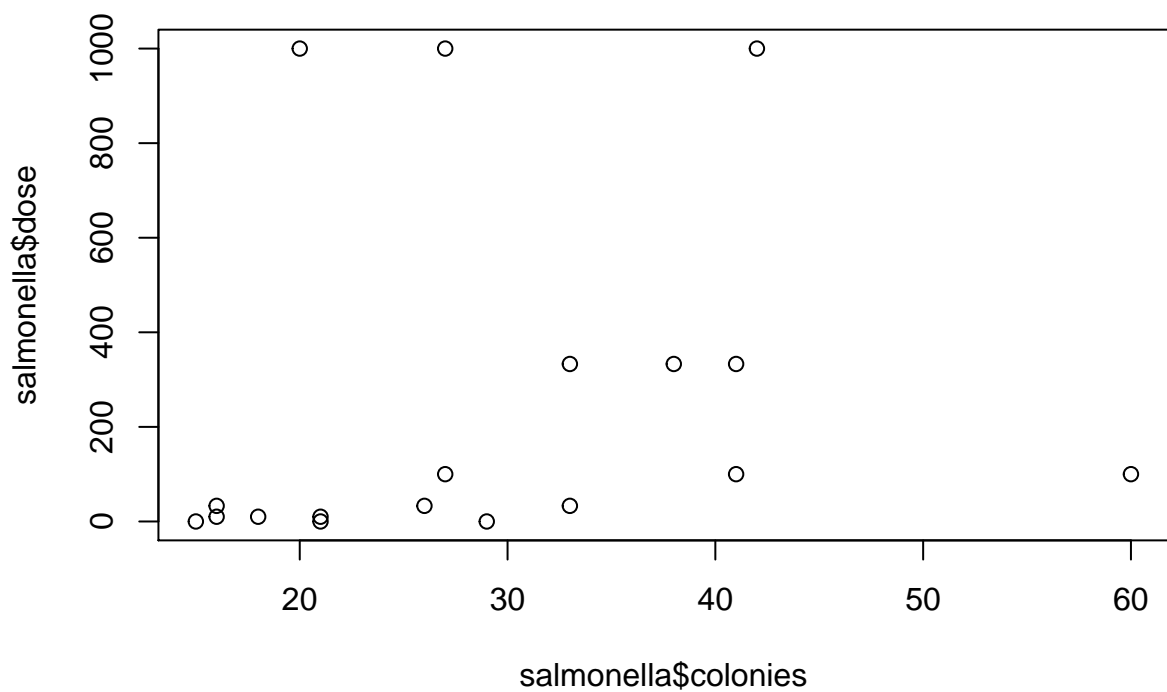
h)



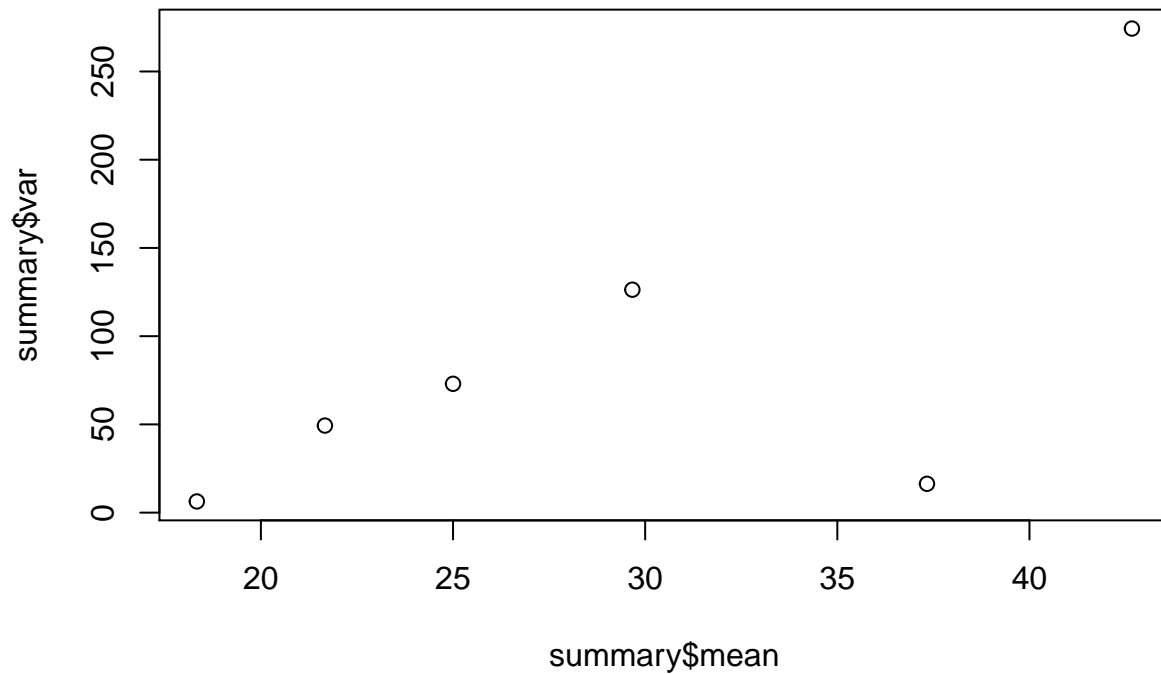
There are a few points (one in particular) with especially high leverages and large negative residuals.

Chapter 5 Problem 2, page 99

a)



The scatter plot and bar plot of means by dose shows that in general the higher the dose the more colonies. A dose of 100 produces the most colonies. After going above a dose of 100, the number of colonies decrease.



There is high variability for doses with a very high mean so there may be overdispersion, and an overdispersion parameter would be needed to account for this.

c)

```
##
## Call:
## glm(formula = colonies ~ dose, family = "poisson", data = salmonella.2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5737  -0.6820  -0.1110   0.6041   2.4989
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.0758     0.1240  24.798 < 2e-16 ***
## dose10        -0.1671     0.1832  -0.912  0.361869
## dose33         0.1431     0.1695   0.844  0.398427
## dose100        0.6776     0.1523   4.449 8.62e-06 ***
## dose333        0.5441     0.1559   3.490 0.000484 ***
## dose1000       0.3142     0.1632   1.926 0.054099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 78.358 on 17 degrees of freedom
## Residual deviance: 33.496 on 12 degrees of freedom
## AIC: 138.03
##
## Number of Fisher Scoring iterations: 4
## [1] 33.49596
```

Yes you could do a log transformation and do standard linear regression as opposed to poisson. This will be done below.

```
##
## Call:
## lm(formula = log(colonies) ~ dose, data = salmonella.2)
##
## Residuals:
```

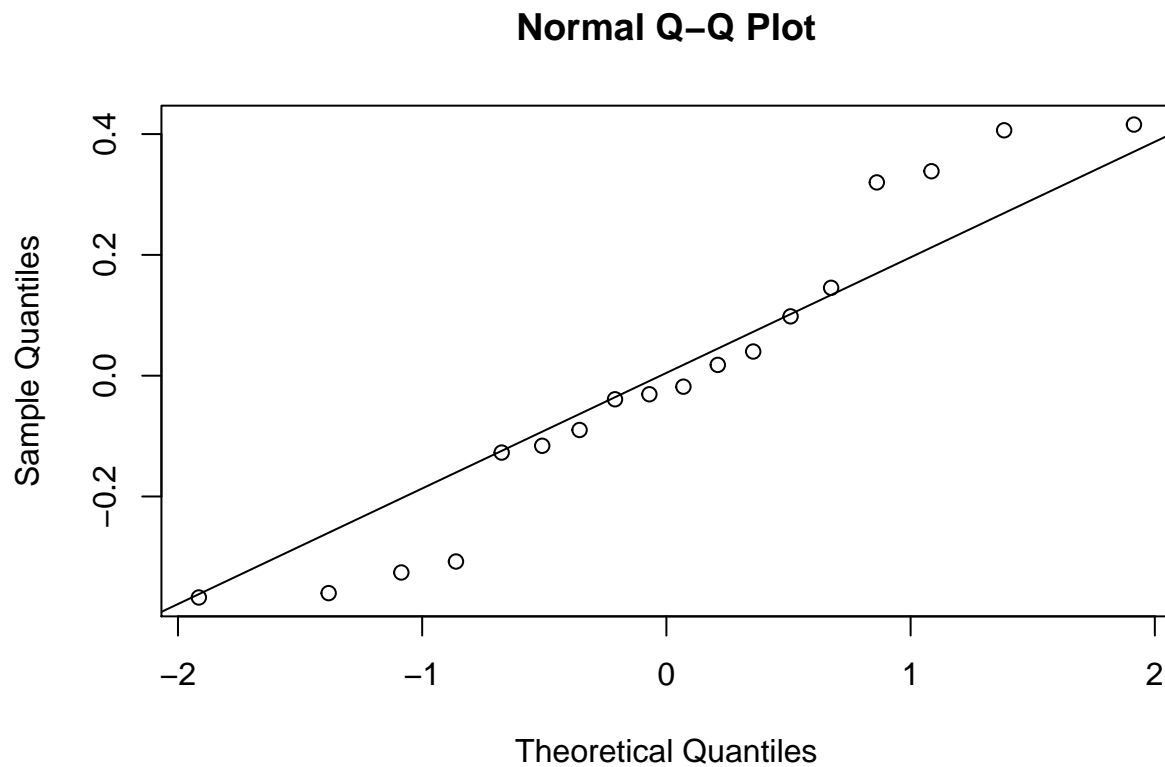
	Min	1Q	Median	3Q	Max
	-0.40541	-0.12727	0.00844	0.13094	0.39459

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.0400	0.1786	17.019	9.1e-10 ***
dose10	-0.1375	0.2526	-0.544	0.5963
dose33	0.1358	0.2526	0.537	0.6008
dose100	0.6613	0.2526	2.618	0.0225 *
dose333	0.5759	0.2526	2.280	0.0417 *
dose1000	0.3031	0.2526	1.200	0.2533

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3094 on 12 degrees of freedom
## Multiple R-squared:  0.5681, Adjusted R-squared:  0.3882
## F-statistic: 3.157 on 5 and 12 DF, p-value: 0.04769
```

d)



The residuals appear to be larger in the upper quantiles, and to some extent the lower quantiles.

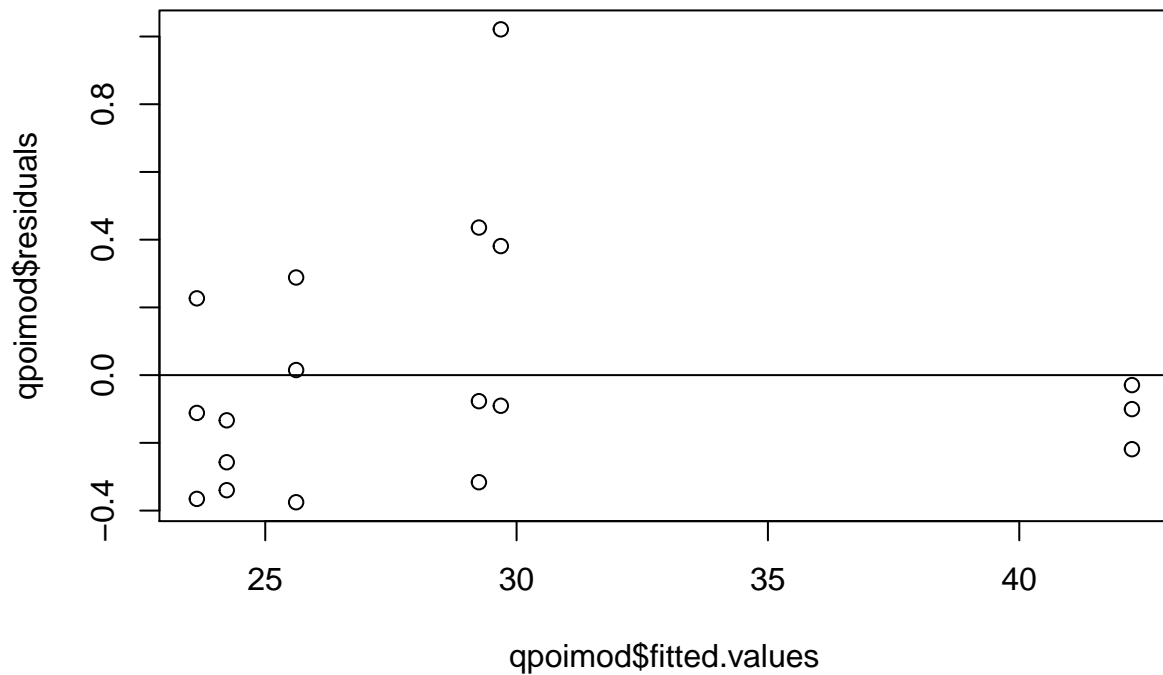
e)

```
##
## Call:
## glm(formula = colonies ~ dose + I(dose^2), family = "quasipoisson",
##      data = salmonella)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0422  -1.4412  -0.5271   0.8173   4.8797
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.163e+00  1.361e-01  23.237 3.55e-13 ***
## dose         2.507e-03  1.040e-03   2.410  0.0293 *
## I(dose^2)    -2.294e-06  1.003e-06  -2.288  0.0371 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.126227)
##
```

```
## Null deviance: 78.358 on 17 degrees of freedom
## Residual deviance: 55.535 on 15 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
## [1] 55.53472
```

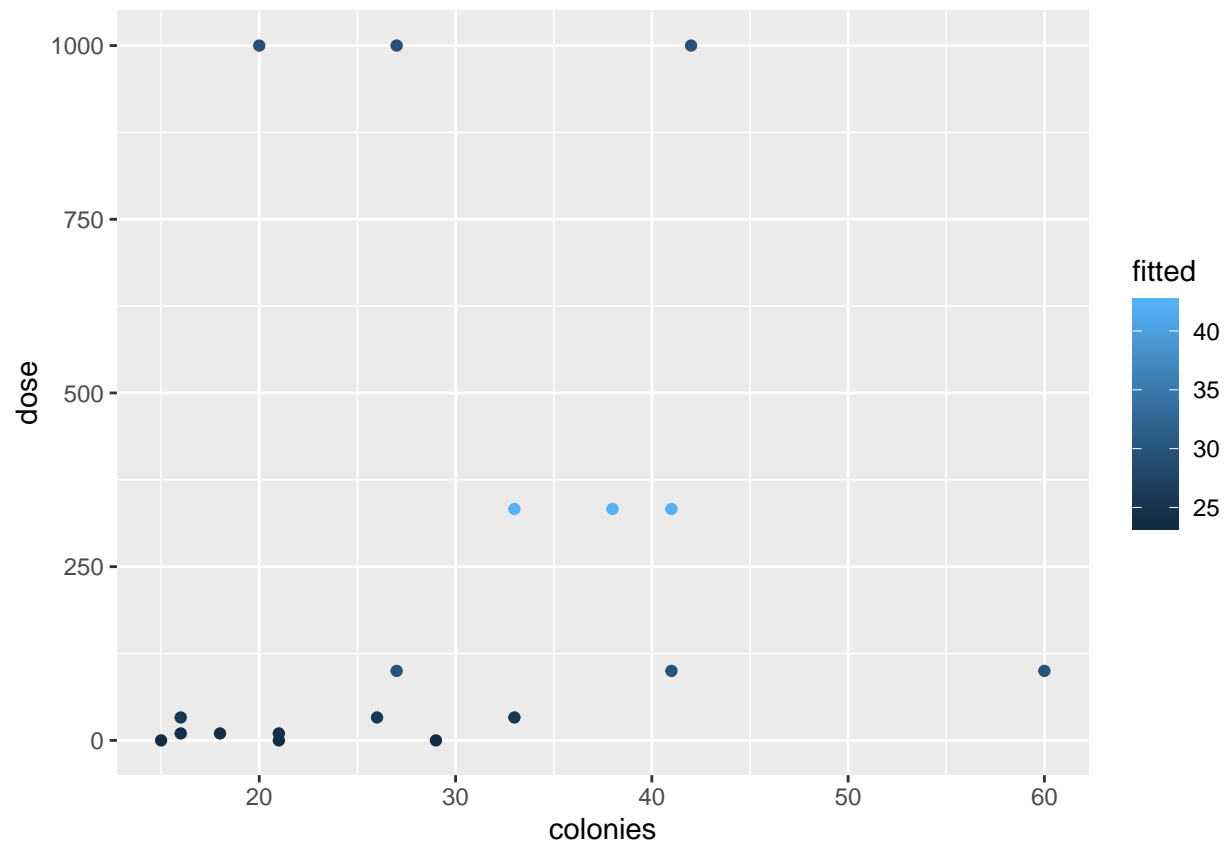
Yes we can determine the deviance, it is shown above. As shown below, the residuals are fairly normal so it is a fairly good fit.

f)



The residuals appear to be mostly normally distributed across the range of fitted values although there is one especially large positive residual

g)



h)

```
##          1          1
## 3.438811 4.247112
```

The predicted mean colonies given a dose of 500 would be 3.8429616 and a 95% confidence interval for the mean would be 3.438811, 4.2471121

i)

The maximum predicted response occurs at a dose of 333