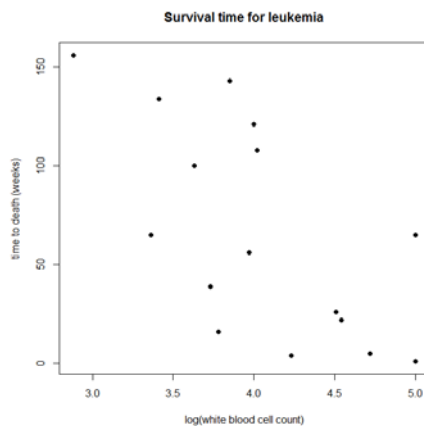Q1

a) Survival time decreases as white blood cell count increases

b) Link function is log

c) Using integration by parts:

$$E[X] = \int x\, \lambda e^{-\lambda x}dx = -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x} \Big|_0^\infty = \frac{1}{\lambda}$$

$$E[X^2] = \int x^2 \lambda e^{-\lambda x}dx = -x^2 e^{-\lambda x} + 2\int xe^{-\lambda x}dx = \frac{2}{\lambda^2}$$

$$Var[X] = E[X^2] - E[X]^2 = \frac{1}{\lambda^2}$$

d) Two models are fitted below :

1. the first one is LM with y=log(survival time)

   lm1 <- lm(log_time ~ log_wbc)

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.0738     2.0393   5.430 6.96e-05 ***
log_wbc        -1.8829     0.4925  -3.823  0.00166 **
```

2. the second is GLM() with y=time

   glm1 <- glm(time ~ log_wbc, family="poisson")

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    7.99867    0.20210   39.58  <2e-16 ***
log_wbc       -0.98639    0.05334  -18.49  <2e-16 ***
```

e) we can see from the residual plot below that the residuals are small, except for the last observation $(x= 5, y= 65, r = 2.027)$. So the model fits the data well

Observations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
x	3.36	2.88	3.63	3.41	3.78	4.02	4	4.23	3.73	3.85	3.97	4.51	4.54	5	5	4.72	5	
y	65	156	100	134	16	108	121	4	39	143	56	26	22	1	1	5	65	
fitted_values	108.25	173.79	82.94	103.04	71.53	56.45	57.58	45.89	75.15	66.76	59.31	34.82	33.80	21.47	21.47	28.30	21.47	
residuals	-4.49	-1.37	1.81	2.91	-7.95	6.09	7.27	-8.02	-4.60	8.09	-0.43	-1.56	-2.17	-5.90	-5.90	-5.41	7.55	
standard_residuals	-0.40	-0.10	0.21	0.30	-0.78	0.91	1.10	-0.91	-0.48	1.14	-0.06	-0.25	-0.35	-0.95	-0.95	-0.82	2.03	

residual plot

Survival time for leukemia (red=lm, blue=glm)
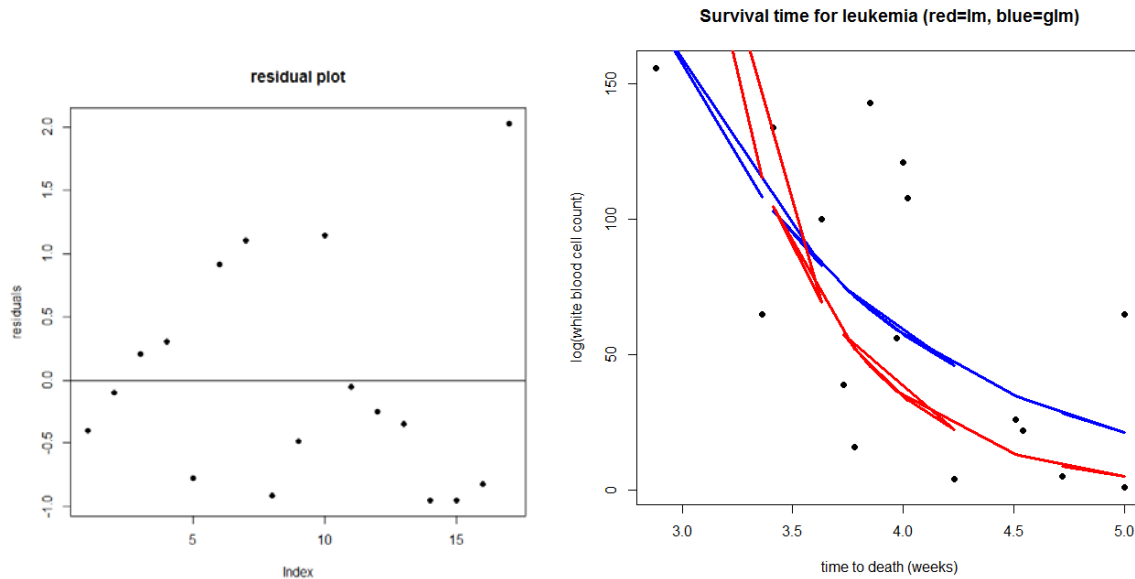
Appendix (R Code)

```
require(dobson)

require(ggplot2)


#preparation of data

leukemia <- dobson::leukemia

log_wbc = leukemia$wbc

time = leukemia$time

log_time = log(leukemia$time)


#plot

plot(y=leukemia$time, x=leukemia$wbc, pch=19, col= "black",
```

```r
    main ="Survival time for leukemia", ylab ="time to death (weeks)", xlab = "log(white blood
cell count)")


#models

lm1 <- lm(log_time ~ log_wbc)

glm1 <- glm(time ~ log_wbc, family="poisson")


#table

table1 <- rbind( x=log_wbc, y= glm1$y, fitted_values= glm1$fitted.values,

            residuals = resid(glm1), standard_residuals = glm1$residuals)

write.csv(x= table1, file = "./York Statistics/Math 6622 - GLM/A2/table.csv" )


#plots

plot(y=leukemia$time, x=leukemia$wbc, pch=19, col= "black",

    main ="Survival time for leukemia (red=lm, blue=glm)", xlab ="time to death (weeks)", ylab
= "log(white blood cell count)")

lines(leukemia$wbc, glm1$fitted, col="blue",lwd=3)

lines(leukemia$wbc, exp(lm1$fitted.values), col="red",lwd=3)


#residual plot

plot(glm1$residuals , pch=19, ylab = "residuals", main = "residual plot")

abline(h = 0)

plot( glm1$fitted , resid(glm1), pch=19)
```