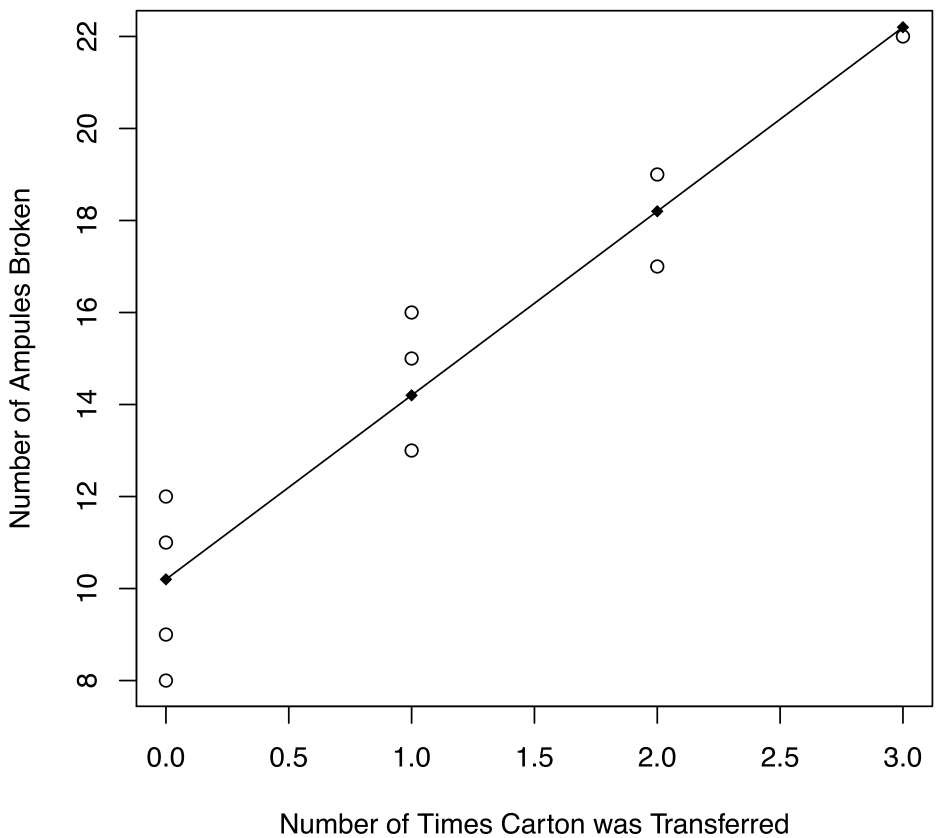
Question 1

Part 1

a) The fitted model for a traditional linear regression is

$$\hat{Y}_i = 10.2 + 4X_i$$



The model is a good fit. Using p-value for the F-statistics, we reject the hypothesis that the regression coefficients are insignificant. Also, the model explains 90% of the variation in the number of ampules broken upon arrival.

b) The expected number of broken ampules are:

X	0	1	2	3
$\hat{Y}$	10.2	14.2	18.2	22.2

c) The estimated increased in the expected number of broken ampules when there are 2 transfers as compared to 1 transfer is 4.

Part 2

a) The fitted Poisson regression model is

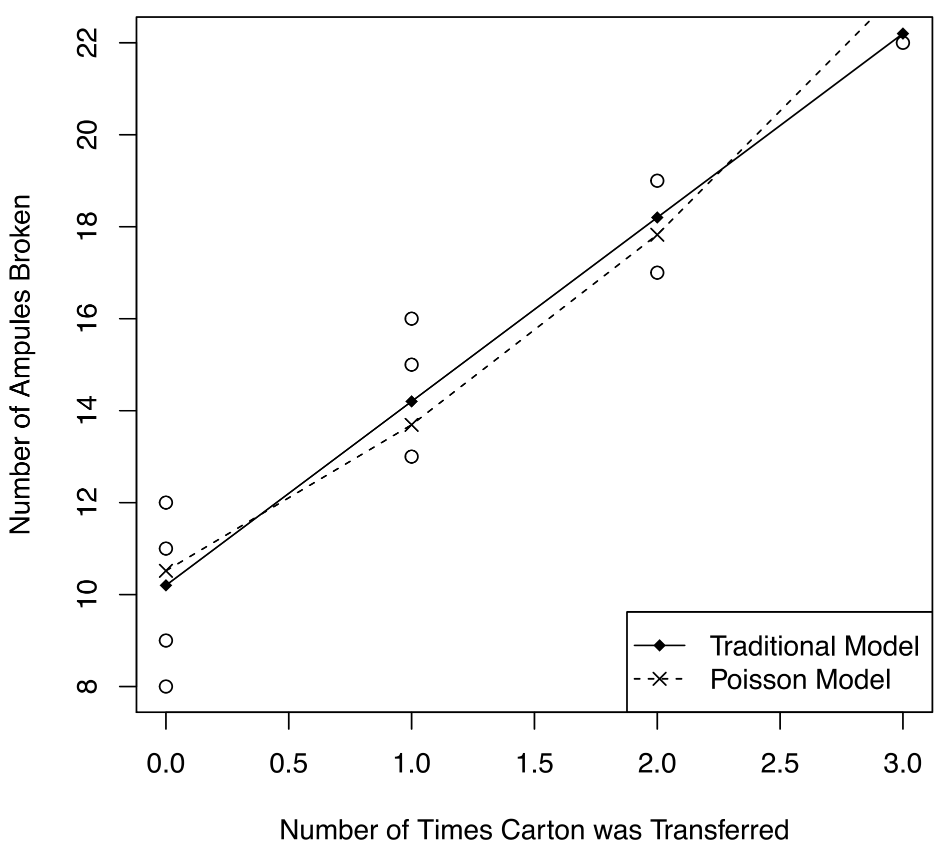
$$\ln(\hat{\lambda}_i) = 2.3529 + 0.2638X_i$$

$$\hat{\lambda}_i = e^{2.3529+0.2638X_i}$$

b) The expected number of broken ampules are:

X	0	1	2	3
$\hat{Y}$	10.52	13.69	17.82	23.21

As compared to the results from the traditional model, the expected values are not much different.



c)

The traditional linear regression model appears to be a better fit to the data than the Poisson regression model.

d) When there are no transfers ($X_i = 0$)

$$\hat{\lambda}_i = e^{2.3529} = 10.52,$$

thefore, the probability that 10 or fewer ampules are broken will be

$$P(Y_i \leq 10) = \sum_{y=0}^{10} \frac{e^{-10.52}(10.52)^y}{y!} = 0.5188$$

e) 95% confidence interval for the slope β_1 is

$$0.1073 \leq \beta_1 \leq 0.4183$$

If the number of times a carton is transferred increases by 1, the expected number of ampules broken upon arrival will increase by about 30% ($e^{0.2638} = 1.3019$) and the log of expected number of ampules broken upon arrival will be between 0.1073 and 0.4183.

Question 2

a)

$$f(y) = \lambda e^{-\lambda y} = \exp(\ln \lambda - \lambda y) = \exp(-\lambda y + \ln \lambda)$$

$$\theta = -\lambda, \quad \phi = 1, \quad a(\phi) = 1, \quad b(\theta) = -\ln \lambda = -\ln(-\theta), \quad c(y, \phi) = 0$$

b)

$$\mathrm{E}(Y) = b'(\theta) = -\frac{1}{\theta} = \frac{1}{\lambda} = \mu$$

The canonical link is

$$\theta = -\lambda = -\frac{1}{\mu}$$

The variance function is

$$\mathrm{Var}(Y) = b''(\theta)a(\phi) = \frac{1}{\theta^2} = \frac{1}{\lambda^2}$$

c) Using the canonical link as a response in a linear regression may results in negative mean responses.

d) Deviance formula for exponential distribution:

$$D(y_i, \hat{\mu}_i) = 2\left[\ln L(\boldsymbol{y}, \phi|\boldsymbol{y}) - \ln L(\hat{\boldsymbol{\mu}}, \phi|\boldsymbol{y})\right]$$

$$
\begin{aligned}
&= 2\left[\ln\prod_{i=1}^{n}\exp\left(\ln\left(\frac{1}{y_i}\right)-\frac{y_i}{y_i}\right)-\ln\prod_{i=1}^{n}\exp\left(\ln\left(\frac{1}{\hat{\mu}_i}\right)-\frac{y_i}{\hat{\mu}_i}\right)\right]\\
&= 2\sum_{i=1}^{n}\left(\ln\frac{1}{y_i}-\ln\frac{1}{\mu_i}-1+\frac{y_i}{\mu_i}\right)\\
&= 2\sum_{i=1}^{n}\left(\frac{y_i-\hat{\mu}_i}{\hat{\mu}_i}-\ln\frac{y_i}{\hat{\mu}_i}\right)
\end{aligned}
$$