Question 13

a) I read in the dataset, and fitted a glm with family = Poisson(). Here is the summary:

```
Call:
glm(formula = sat ~ weight, family = poisson(), data = crabs)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.42841    0.17893  -2.394   0.0167 *
weight       0.58930    0.06502   9.064   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 560.87  on 171  degrees of freedom
AIC: 920.16

Number of Fisher Scoring iterations: 5
```

b) I used the model to predict a female crab's satellites at weight=2.44

```
# b)
pred = predict(p_model, data.frame(weight=2.44), type="response")
print('Predicted Y:')
print(pred)
```

```
[1] "Predicted Y:"
       1
 2.74422
```

c) After extracting the coefficient of B_hat from the model, I used a confidence interval to get the appropriate confidence for B at 95%. I then took 10^value, as the Poisson uses log-linear, so to invert the log I applied an exponential. The 95% confidence interval shows a .46 - .71 multiplicative increase in 1kg of x, or a satellite gain of 2.88 to 5.18 crabs

```
[1] "B_hat:"
   weight
0.5893041
Waiting for profiling to be done...
    2.5 %    97.5 %
0.4597002 0.7144983
    2.5 %
2.882041
   97.5 %
5.182011
```

Question 14

a) I imported the MASS library and used the glm.nb function to produce this summary: The Dispersion parameter is .931, so it suggests a slight under-dispersion. It will fit just a little bit better, its AIC is also slightly lower.

```
Call:
glm.nb(formula = sat ~ weight, data = crabs, init.theta = 0.9310592338,
    link = log)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8647     0.4048  -2.136   0.0327 *
weight        0.7603     0.1578   4.817 1.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.9311) family taken to be 1)

    Null deviance: 216.43  on 172  degrees of freedom
Residual deviance: 196.16  on 171  degrees of freedom
AIC: 754.64

Number of Fisher Scoring iterations: 1


              Theta:  0.931
          Std. Err.:  0.168

 2 x log-likelihood:  -748.644
```

b) The confidence interval is wider with the negative binomial because it accounts for the under-dispersion, meaning that it takes into consideration the variance of the model, which Is more flexible when using a Negative Binomial dist.

```
[1] "B_hat2:"
   weight
0.7602787
Waiting for profiling to be done...
    2.5 %    97.5 %
0.4207032 1.1136662
```