Problem 1

(a) The prediction equation is

$$\hat{\pi}_1(x) = 0.0023823 + 0.0016712x.$$

(i) $\hat{\pi}_1(0) = 0.002382292$ and $\hat{\pi}_1(7) = 0.014080648$.

(ii) $\frac{\hat{\pi}_1(0)}{\pi(0)} = 0.8493864$ and $\frac{\hat{\pi}_1(7)}{\pi(7)} = 0.5350646$.

(b) The prediction equation is

$$\hat{\pi}_2(x) = 0.0026353 + 0.0006682x.$$

(i) $\hat{\pi}_2(0) = 0.002635315$ and $\hat{\pi}_2(7) = 0.007312809$.

(ii) $\frac{\hat{\pi}_2(0)}{\hat{\pi}_1(0)} = 1.10621$ and $\frac{\hat{\pi}_2(7)}{\hat{\pi}_1(7)} = 0.5193517$.

Since the probability $\hat{\pi}_2(7)$ is just half of that we estimated in (a), we can conclude that the result is sensitive to this single malformation.

(c) The prediction equation is

$$\hat{\pi}_3(x) = 0.0025032 + 0.0006895x.$$

(i) $\hat{\pi}_3(0) = 0.002503196$ and $\hat{\pi}_3(4) = 0.005261370$.

(ii) $\frac{\hat{\pi}_3(0)}{\hat{\pi}_1(0)} = 1.050751$ and $\frac{\hat{\pi}_3(4)}{\hat{\pi}_1(4)} = 0.3736597$.

Since the probability $\hat{\pi}_3(4)$ is just 37% of that we estimated in (a), we can conclude that the result is sensitive to choice of scores.

(d) We are going to fit a logistic regression model. The prediction equation is

$$logit(\hat{\pi})_4(x) = -5.9605 + 0.3166x.$$

Interpret: When alcohol score is increased by 1 unit, the odds of present are multipled by $e^{0.3166} = 1.372453$.

Problem 2

(a) The linear model is

$$Y = -0.1448709 + 0.0003227 * weight.$$

Interpret: The intercept corresponds to the estimated probability of a satellite at $x = 0$. However, it is negative, which means it is not a probability. The slope represents an increase in the probabilty of haveing a sastellite when weight is increased by one gram, which is 0.0003227.

The predicted probability at the highest observed weight of 5.20kg is

$$Y(5200) = -0.1448709 + 0.0003227 * 5200 = 1.533186.$$

It is not a probability at all.

The linear regression shos that using MLE will fail.

(b) The logistic model is

$$logit(\hat{\pi}) = -3.6947264 + 0.0018151 * weight.$$

The fitted value at a weight of 5.20 kg is $\hat{\pi} = 0.9968$, and $logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1-\hat{\pi}}) = 5.74$.

Problem 3

(a) The probit model is

$$probit(\hat{\pi}) = -2.2382945 + 0.0010990 * weight.$$

(b) The fitted value at the highest obseved weight 5.20 kg is 0.9997462.

(c) The difference between the $\hat{\pi}$ values at the upper and lower quartiles of weight is

$$0.8143 - 0.4839 = 0.3304.$$

(d) The weight of crabs has a normal distribution $weight \sim N(-\frac{\alpha}{\beta}, \frac{1}{\beta^2})$. It means that if the crab weight is lighter than $-\frac{\alpha}{\beta} = 2.04kg$, it tends to have no satellite, vice verse.

For a crab that is about one standard deviation $\frac{1}{\beta^2} = 910g$, the probability of havng a satellite is .841. For 2 and 3 standard deviation, the probability will be .97 and .99 respectively.

Problem 4

(a) $logit(\hat{\pi}) = -3.5561 + 0.0532 * Inc$

(b) Since $\beta > 0$, the probability of one possesses credit card is increased by income increasing.

(c) $logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1-\hat{\pi}}) = 0$.

when income = 66.86, using the prediction equation above, we get

$$\hat{\pi} = \frac{exp(-3.5561 + 0.0532 * 66.86)}{1 + exp(-3.5561 + 0.0532 * 66.86)} \approx 0.50.$$

Problem 5

(a) The prediction equation is $log(\hat{\mu}(x)) = -0.4284053 + 0.0005893x$.

(b) We estimate the log of the mean weight to be

$$-0.4284053 + 0.0005893 * 2440 = 1.009487.$$

therefore the estimated mean weight is $exp(1.009487) = 2.744193$ satellites.

(c) An increase of 1 gram increases the estimated mean number of satellites of $e^{0.000589} = 1.000589$.

From R,the 95% CI for β is given by

$$(0.0004597002, 0.0007144983).$$

A CI for the multiplicative effect e^{β} is given by

$$(1.0004598, 1.0007148).$$

(d) If the mean of Y is independent of the weight, it means that $\beta = 0$. The z value is 9.064 which is significantly different from 0. Therefore we strongly reject the null hypothesis that satellite count is independent of the weight of the crab.

(e) According to ANOVA table in R, the chi-squred statistic on 1 d.f. is 71.925 and its p-value is really small. So weight is highly significant.