Logistic Regression

1. The NFL collected data on field goal success in professional American Football games. The data consisted of a binary variable for success (Success=1, Failure=0) and the distance in yards from the goal. The data were submitted to a logistic regression, a portion of the R out of the analysis is given below

```
Call: glm(formula = Success ~ Yards, family = binomial("logit"),
data = fieldgoal)


Deviance Residuals:
    Min       1Q     Median     3Q      Max
 -2.6568  0.2718   0.4166   0.6938   1.4750
Coefficients:
              Estimate   Std. Error
 (Intercept)   5.69788    0.45110
 Yards        -0.10991    0.01058
```

a) Write the formula of the logistic regression, explaining the terms used in the expression.

(3 marks)

b) Calculate the z-value of the estimates

(5 marks)

c) What are the estimated odds that a 60 yard field goal will succeed

(4 marks)

d) What is the probability that a 30 yard field goal will succeed, and what is probability it will fail.

(5 marks)

2. Sahoo and Pandalai (1999) conducted a study on the success or failure of finding gold deposits as a function of water/chemical factors: As level, Sb level, and presence==1 or absence==0 of lineament. A portion of the R analysis is given below

```
Call: glm(formula = Presence ~ Aslevel+Sblevel+LineamentProx,
family = binomial("logit"), data = gold)


Deviance Residuals:
    Min       1Q     Median      3Q      Max
 -2.6568   0.2718   0.4166   0.6938   1.4750
Coefficients:
                   Estimate   Std. Error
 (Intercept)       -7.6096      3.1661
 Aslevel            1.2046      0.4899
 Sblevel            1.4210      0.7301
 LineamentProx      3.1973      1.8911
```

a) Write the formula of the logistic regression, explaining the terms used in the expression.

(3 marks)

b) Calculate the z-value of the estimates.

(7 marks)

c) What are the probability of finding a gold deposit given an As level of 3, Sblevel of 4 and no lineament.

(4 marks)

d) Given a 95% rejection criteria with a z-criteria of (-1.96, 196), are the estimates of the coefficents significant.

(3 marks)

Formulae

Counting Rules

Permutations

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

Combinations

$$^{n}C_{r} = \begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{r!(n-r)!}$$

Addition of Probability

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Conditional Probability

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Bayes Formula

$$p(H_1|A) = \frac{p(A|H_1)p(H_1)}{p(A|H_1)p(H_1) + p(A|H_2)p(H_2)}$$

Probability Mass Functions

$$E[X] = \Sigma_{i=1} x_i p(x_i)$$

$$Var[X] = \Sigma_{i=1} (x_i - \mu)^2 p(x_i)$$

Geometric Distributions

$$p(k) = q^{(k-1)}p, \quad k = 1, 2, \ldots$$

$$E[k] = \frac{1}{p} \quad Var[k] = \frac{q}{p^2}$$

Binomial Distributions

$$p(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots n$$

$$E[k] = np \quad Var[k] = npq$$

Poisson Distributions

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

$$E[k] = \lambda \quad Var[k] = \lambda$$

Chi squared

$$\chi_{GoF} = \sum \frac{(O-E)^2}{E} \; \chi^2_{k-1}, \;\; \chi_{Ind} = \sum \frac{(O-E)^2}{E} \; \chi^2_{(r-1)(c-1)}$$

Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots$$

$$y_i = \beta_0 + \beta_1 x_{i1}$$

$$\beta_1 = \frac{Cov(x,y)}{SS_{xx}}, \;\; \beta_0 = \bar{y} - \beta_1 \bar{x}$$

Logistic Regression

$$p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}, \quad \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots$$