Binary Logistic Regression: Tutorial Sheet

1. What is logistic regression? How does it differ from linear regression? Under what circumstances would you use it?

2. What is the Likelihood Ratio Test? Describe how it used in Logistic regression.

3. Suppose the odds of an outcome are 4. What is the probability of that outcome?

4. Suppose the probability of an outcome is 70%. What is the odds of that outcome occuring? Answer this question for probabilities of 40%, 60% and 90% also.

5. The usual assumptions placed on the error terms in ordinary least squares regression are:

   * independently distributed
   * identically distributed (equal variance)
   * normally distributed

   Which of these assumptions are violated when dealing with binary response data? Explain briefly how each is violated.

6. What is a dummy variable? Explain how it is used in Logistic Regression. Support your answer with an example.

7. Suppose that, out of a samle of 100 women and 100 men, 80 men drank alcohol in the last week, while 20 women drank alcohol in past week. Compute the odds ratio for Women to men.

8. What is a logit? how is it computed into a probability?

9. A predictive model was used to predict onset of Coronary Heart Disease, based on a set of medical measurements and characteristics. Comment on the significance of each of the predictor variables used in this model.

   sbp systolic blood pressure

   tobacco cumulative tobacco (kg)

   ldl low density lipoprotein cholesterol

   adiposity A BMI index

   famhist family history of heart disease, a factor with levels (Absent/Present)

   typea type-A behavior

   obesity A BMI index

   alcohol current alcohol consumption

   age age at onset

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)    -6.1507209  1.3082600  -4.701 2.58e-06 ***
sbp             0.0065040  0.0057304   1.135 0.256374
tobacco         0.0793764  0.0266028   2.984 0.002847 **
ldl             0.1739239  0.0596617   2.915 0.003555 **
adiposity       0.0185866  0.0292894   0.635 0.525700
famhistPresent  0.9253704  0.2278940   4.061 4.90e-05 ***
typea           0.0395950  0.0123202   3.214 0.001310 **
obesity        -0.0629099  0.0442477  -1.422 0.155095
alcohol         0.0001217  0.0044832   0.027 0.978350
age             0.0452253  0.0121298   3.728 0.000193 ***
```

10. *This question is a continuation of the previous question.* Using the following binary logistic model, which is used to predict the onset of Coronary Heart Disease, compute the predicted outcomes for the cases below;

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.334452   0.897809  -7.055 1.72e-12 ***
age          0.055040   0.009948   5.533 3.15e-08 ***
ldl          0.179891   0.055027   3.269  0.00108 **
typea        0.037914   0.011885   3.190  0.00142 **
tobacco      0.075031   0.025699   2.920  0.00350 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

(i) 52 year old person, ldl = 5.73 , typea = 49, tobacco = 12

(ii) 63 year old person, ldl = 4.41 , typea = 55 , tobacco = 0.01

11. There are three variants for the forward selection procedures used by SPSS. Name these three.