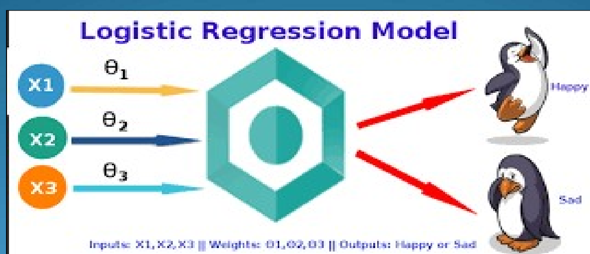
Logistic Regression



When And Why

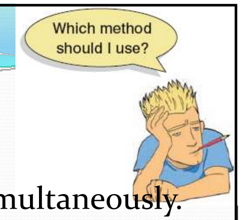
- To predict an outcome variable that is categorical from one or more categorical or continuous predictor variables.
- Used because having a categorical outcome variable violates the assumption of linearity in normal regression.

Slide 2

Aims

- When and Why do we Use Logistic Regression?
  - Binary
  - Multinomial
- Theory Behind Logistic Regression
  - Assessing the Model
  - Assessing predictors
  - Things that can go Wrong
- Interpreting Logistic Regression

Methods of Regression



- Forced Entry: All variables entered simultaneously.
- Hierarchical: Variables entered in blocks.
  - Blocks should be based on past research, or theory being tested. Good Method.
- Stepwise: Variables entered on the basis of statistical criteria (i.e. relative contribution to predicting outcome).
  - Should be used only for exploratory analysis.

With One Predictor

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_1+\varepsilon_i)}}$$

- Outcome
  - We predict the *probability* of the outcome occurring
- *b_o and b_o*
  - Can be thought of in much the same way as multiple regression
  - Note the normal regression equation forms part of the logistic regression equation

Slide 5

With Several Predictor

$$P(Y) = \frac{1}{1+e^{(b_0+b_1X_1+b_2X_2+\dots+b_nX_n+\varepsilon_i)}}$$

- Outcome
  - We still predict the *probability* of the outcome occurring
- Differences
  - Note the multiple regression equation forms part of the logistic regression equation
  - This part of the equation expands to accommodate additional predictors

Slide 6

Binary classification performance metrics

- Confusion matrix
- Accuracy

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy, Precision , Recall,F1 Score

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = 2\frac{PR}{P + R}$$