### R Squared

1. variation in the o/p variable explained by the model

2. Unexplained variation

Y = b0 + b1X + e

Variation in Y Variation in Y explained by the model Variation in Y not explained by the model

SST = SSR + SSE

sum of squares of total variation = sum of squares of explained variation + sum of squares of error

R Squared = SSR/SST = 1 - SSE/SST

—------------------------------------------------------------------------------------------------------------

### Deriving logit function

1. **p >= 0**
2. **p <= 1**

p = e^(b0 + b1X) - (1)

p = e^(b0 + b1X) / e^(b0 + b1X) + 1 -(2)

z = b0 + b1X

p = e^z / e^z + 1 -(3)

Logit function

q = 1-p

q = 1- e^z / e^z + 1 -(4)

(3)/(4)

p/1-p = e^z

Log on both sides,

log(p/1-p) = z

log(p/1-p) = b0 +b1X

—------------------------------------------------------------------------------------------------------------------

### Model performance

Accuracy => Train = 95%

**Accuracy => Test = 31%**

**Overfitting**

**Low BIAS and High VARIANCE**

**Decision trees**

Feature selection - consider fewer feature combinations

—------------------------------

Accuracy => Train = 39%

**Accuracy => Test = 37%**

**Underfitting**

**High BIAS and low VARIANCE**

**Linear models**

**Model is too simple**

**—----------------------------------------------------------------------------------------------------------------------------**