# MATHS 7107
# Data Taming
# Week 8

Shenal Dedduwakumara

School of Mathematical Sciences, University of Adelaide

# Genaralized Linear Models (GLMs)

- It is common to find response variables which do not fit the standard assumptions of the linear models (normally distributed errors, constant variance, etc.), for example: count data, dichotomous variables, truncated data, etc.
- Generalized linear models (GLMs) expand the the well known linear model to accommodate non-normal response variables in a single unified approach.
  *http://halweb.uc3m.es/esp/Personal/personas/durban/esp/web/notes/glm.pdf*

# Example: Ashes data

```
library(tidyverse)
ashes_avg<- read_csv("ashes_avg.csv")
head(ashes_avg)
```

```
## # A tibble: 6 x 3
##   batter   role    avg
##   <chr>    <chr> <dbl>
## 1 Anderson bowl   3.5
## 2 Archer   bowl   6.86
## 3 Bancroft bat   11
## 4 Burns    bat   39
## 5 Buttler  bat   24.7
## 6 Curran   bowl  16
```

# Example: Ashes data

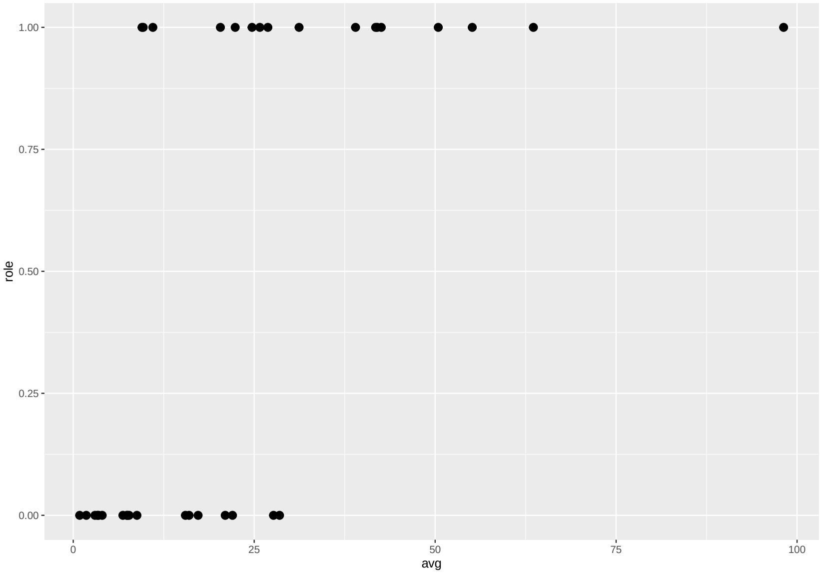
```
ashes_avg %>%
  count(role)
```

```
## # A tibble: 2 x 2
##   role      n
##   <chr> <int>
## 1 bat      17
## 2 bowl     18
```

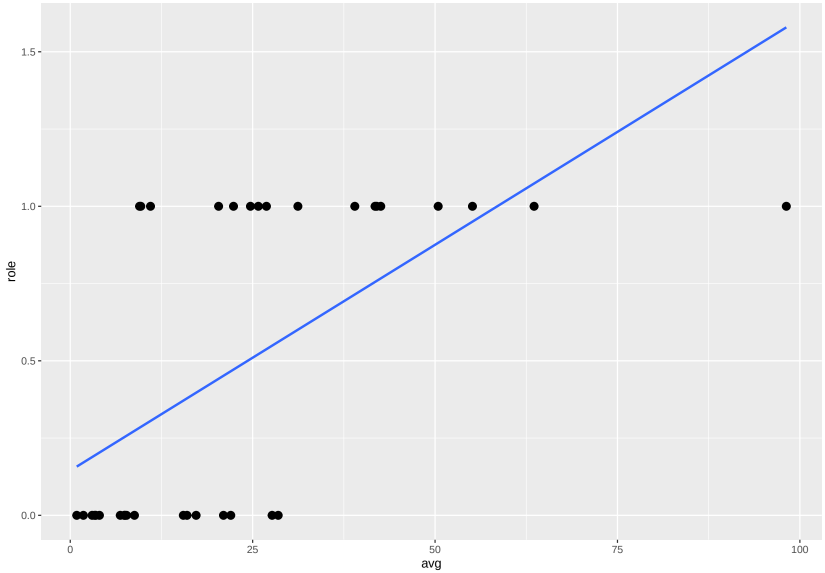
# Example: Ashes data

```
ashes_avg %>%
  mutate(role=recode(role,bat=1,bowl=0)) %>%
  ggplot(aes(avg,role))+
  geom_point(size=3)
```

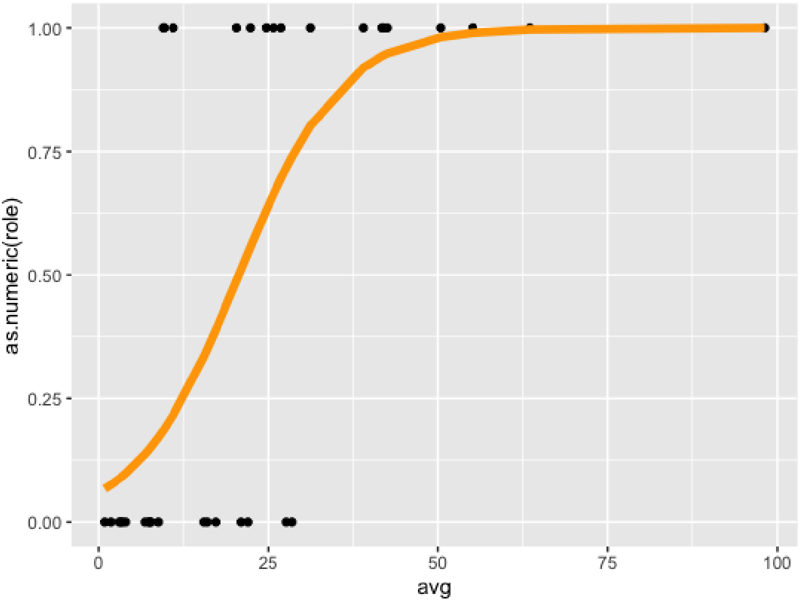
# Example: Ashes data



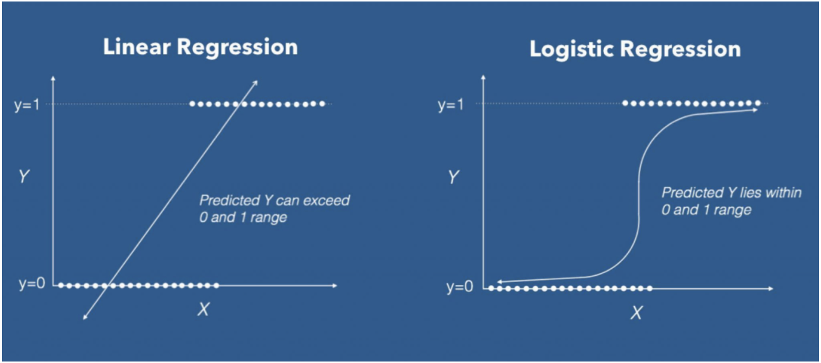
# Example: Ashes data



# Example: Ashes data



# Logistic regression



# Logistic regression

- Response variable are measured on a binary scale
- For example the response may be yes or no; present or absent; dead or alive.

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + ... + \beta_{ik}x_{ik}$$

- For ashes example:

$$p_i = P(Y_i = bat | X_i = avg)$$

# Logistic regression for ashes data

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{i1}x_{i1}$$

# Logistic regression for ashes data

$$p_i \quad ?$$

# Logistic regression for ashes data

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_{i1}x_{i1}$$

$$\left(\frac{p_i}{1-p_i}\right) = e^{\beta_0 + \beta_{i1}x_{i1}}$$

$$p_i = \frac{e^{\beta_0 + \beta_{i1}x_{i1}}}{1 + e^{\beta_0 + \beta_{i1}x_{i1}}}$$