Multiple Regression

We want to build a model with several predictors

- We know that behavior is not determined by one factor, but several at the same time
- Want to examine the effect of a variable on an outcome above and beyond the effect of many others

Multiple Regression: Regression that allows more than one independent, or predictor, variable to help us predict the dependent variable

The goal is to explain more of the variance in the dependent variable, so you want to use many more predictors to account for their effects on the outcome

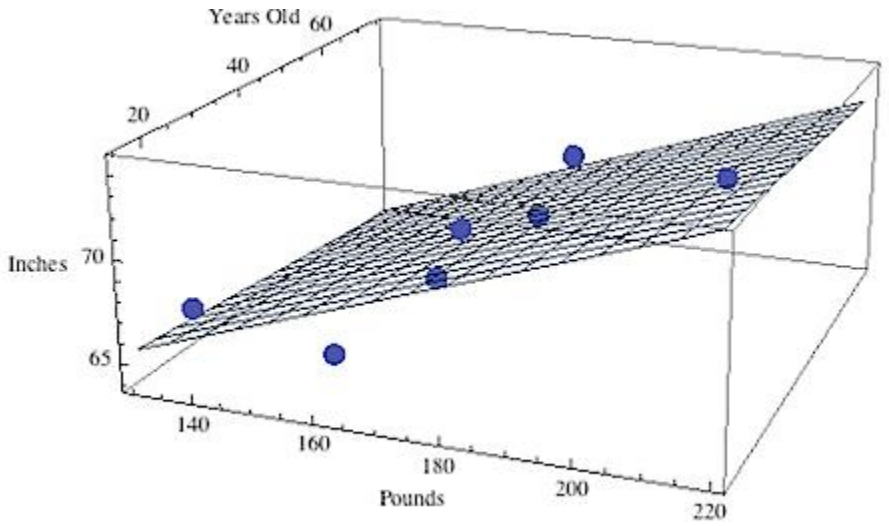
Hyperplane of best fit

We want to build a model with several predictors

- We know that behavior is not determined by one factor, but several at the same time
- Want to examine the effect of a variable on an outcome above and beyond the effect of many others

Multiple Regression: Regression that allows more than one independent, or predictor, variable to help us predict the dependent variable

The goal is to explain more of the variance in the dependent variable, so you want to use many more predictors to account for their effects on the outcome



Steps

- To make predictions in a multiple regression, you:
- Find the values for the intercept and the coefficients
- Enter the values of each X variable, for the case you want to predict
- Multiply and add everything to get the predicted Y
- $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3$

Example: Predict College GPA from:

- Predictors
  - X1 = High School G.P.A
  - X2 = ACT
  - X3 = Number of Extra Curricular activities

- Regression Intercept and Coefficients
  - B0 = -1.92
  - $B_{HS\_GPA} = .36$
  - $B_{ACT} = .13$
  - $B_{ExtraCurriculars} = .10$

- **Prediction: A person with a H.S. GPA of 3.8; ACT of 28, and 4 Extra cirricular activities**
  - **Predicted College G.P.A. = -1.92 + (.36 * 3.8) + (.13*28) + (.10*4)**
  - **Predicted College G.P.A = 3.48**

Multicollinearity

- Because a predictor only has a strong coefficient if it UNIQUELY explains an outcome, **you need to make sure that the predictors do not overlap and correlate highly with another predictor**

- **Multicollinearity:** there is a strong correlation between two or more predictors.

- If there is multicollinearity between predictors, it will appear like the variables do not predict the outcome, when they actually do.

Multicollinearity

Consequences of multicollinearity-

- If you have variables that are multicollinear, **then the coefficients cannot be estimated correctly**

- Example: Predicting height with hand length
  - If you are predicting a person's height with the length of their right hand, there is a correlation
  - However, if you control for the length of the person's left hand, then the length of the right hand doesn't add anything extra
  - If you use the left-hand length to predict a person's height, and control for the length of the right-hand, then the left-hand length doesn't add anything extra

Benefits and Limitations of Regression

Benefits

- Can handle multiple predictors
- Explanation of predictor variable is straightforward
- Fast to compute

Limitations

- Only works for linear relationships
- Addressing non-linearity is difficult
- Prone to overfitting
- Does not handle excess and correlated variables very well
- Only meant for ratio/interval outcomes and requires all data to be standardized.