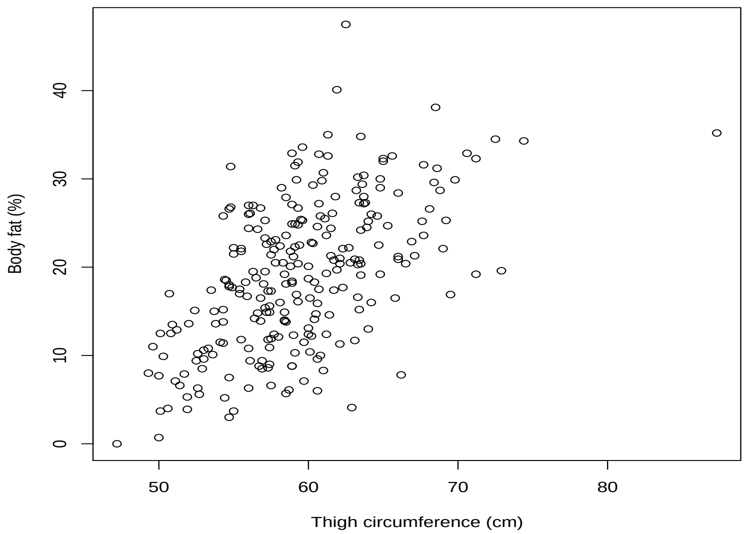
University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

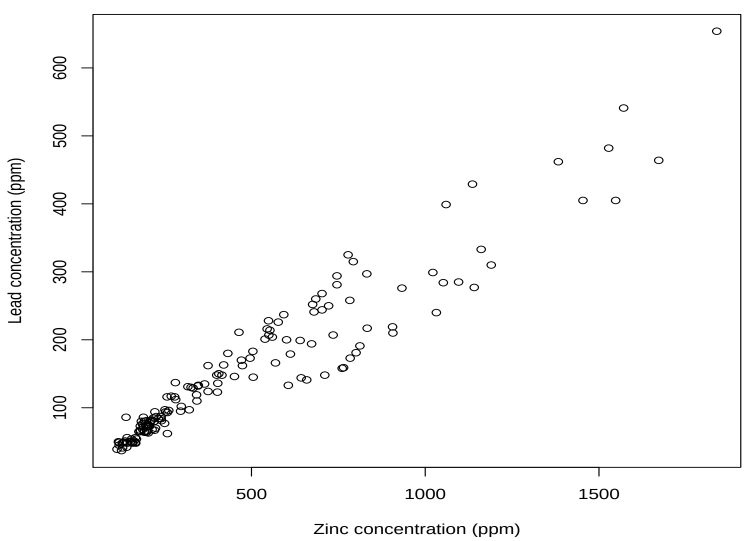
Simple regression analysis

Introduction

Regression analysis is a statistical method used to discover how one variable is associated to another variable. It is useful in explaining the variability and predicting one variable from another variable. Consider the following "scatterplot" of the percentage of body fat against thigh circumference.



Another example: Concentration of lead against the concentration of zinc.



What do you observe?

Is there an equation that can model the picture above?

- Regression model equation:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where

  - y response variable (random)
  - x predictor variable (non-random)
  - β_0 intercept (non-random)
  - β_1 slope (non-random)
  - $\epsilon_1, \ldots, \epsilon_n$ are random error terms (disturbances)
  - Gauss-Markov conditions: $E(\epsilon_i) = 0, var(\epsilon_i) = \sigma^2$, and $\epsilon_1, \ldots, \epsilon_n$ are independent.
  - Distribution assumption: ϵ random error term, $\epsilon \sim N(0, \sigma)$.

- Using the method of least squares we estimate β_0 and β_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$

- The fitted values are computed using the fitted line equation: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and the residuals (difference between the observed y_i value and the fitted value $\hat{y}_i$) are computed using $e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$.

- Useful: The fitted line can be expressed as $\hat{y}_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$.