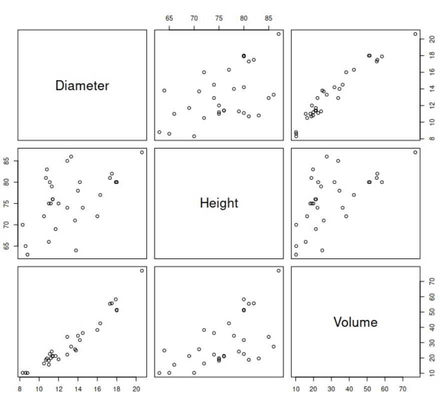
MULTIPLE LINEAR REGRESSION

There are now $p>1$ covariates $x_1, \dots, x_p$.

Example: "trees" R dataset contains data on 31 cherry trees. In particular, we have
- diameter (inches)
- height (feet)
- volume

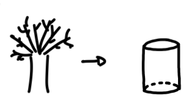
With 3 or more variables we can no longer visualize the relationship with a scatterplot,
We have to use a "matrix of scatterplots" which shows all the PAIRWISE combinations.



The goal is to predict the volume given the other 2 measures

We could use a linear model $\quad volume_i = \beta_1 + \beta_2\, diameter_i + \beta_3\, height_i + \varepsilon_i$

However, in this case, we obtain a better fit if we consider a transformation of the covariates.
If we think at the shape of a tree, we could think of approximating it to a cylinder



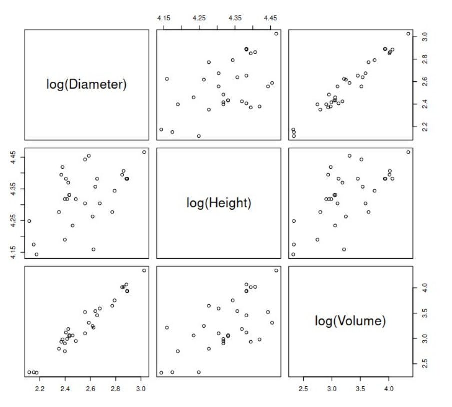
$$volume = \pi \cdot radius^2 \cdot height = \pi \cdot (d/2)^2 \cdot height$$

Hence we could specify a model where $\quad volume_i \approx \pi \cdot \left(\frac{diameter_i}{2}\right)^2 \cdot height_i \qquad \leftarrow$ NOT LINEAR

However, the relationship can be linearized

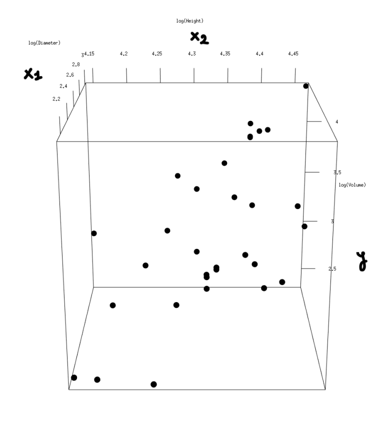
$$\log(volume_i) \approx \log\pi + \log 4 + 2\log(diameter_i) + \log(height_i)$$

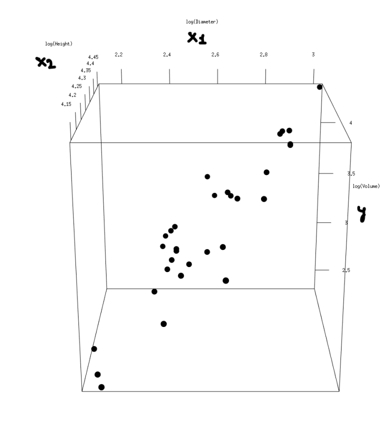
We can consider the transformed variables
- $Y = \log(volume)$
- $x_1 = \log(diameter)$
- $x_2 = \log(height)$



With 2 or more covariates we can not see the JOINT effect they have on y, but only the INDIVIDUAL (marginal) effect of 1 covariate at a time

Only in the case of two covariates we can still see the joint effect using a 3D representation





The goal of the multiple lm is to study the JOINT EFFECT of the covariates on y.

MODEL SPECIFICATION

We now observe $(y_i, x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ip})$ for $i = 1, \dots, n$.

y_i: response variable; $x_{i1}, \dots, x_{ip}$: p covariates

Define the model

$$Y_i = \mu_i + \varepsilon_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, \dots, n$$

If we want to include the intercept, we define $x_{i1} = 1$ for all $i = 1, \dots, n$ (constant variable) and we obtain the model

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, \dots, n$$

NOTATION:

$$\begin{cases} Y_1 = \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_j x_{1j} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ \vdots \\ Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip} + \varepsilon_i \\ \vdots \\ Y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_j x_{nj} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases} \Rightarrow \underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} \quad \underline{x}_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{bmatrix} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

for $j = 1, \dots, p$

$$\Rightarrow \underline{Y} = \beta_1 \underline{x}_1 + \dots + \beta_p \underline{x}_p + \underline{\varepsilon}$$

$$\Rightarrow \underline{Y} = \sum_{j=1}^{p} \beta_j \underline{x}_j + \underline{\varepsilon}$$

$$\Rightarrow \underset{n\times 1}{\underline{Y}} = \underbrace{\underset{n\times p}{X}\ \underset{p\times 1}{\underline{\beta}}}_{\underset{n\times 1}{\underline{\mu}}} + \underset{n\times 1}{\underline{\varepsilon}}$$

where

$$\underset{(n\times p)}{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix} = [\underline{x}_1\ \underline{x}_2\ \dots\ \underline{x}_j\ \dots\ \underline{x}_p] = \begin{bmatrix} \tilde{\underline{x}}_1^T \\ \vdots \\ \tilde{\underline{x}}_i^T \\ \vdots \\ \tilde{\underline{x}}_n^T \end{bmatrix}$$

(row i: $\tilde{\underline{x}}_i^T$; column j: $\underline{x}_j$)

→ $\underline{x}_j$ is the j-th covariate observed on the n units (n-dim vector)

→ $\tilde{\underline{x}}_i$ is the vector of the values of the p covariates on the i-th unit (p-dim vector)

and $\quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$

- $\underline{Y}$ is a vector of random variables $(n\times 1)$
- X is a matrix of known constants $(n\times p)$
- $\underline{\beta}$ is a vector of unknown constants $(p\times 1)$
- $\underline{\varepsilon}$ is a vector of random variables $(n\times 1)$

The assumptions don't change (they are just adjusted for the general case)

1. • normality, homoscedasticity, corr = 0 $\quad \rightarrow \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \quad i = 1, \dots, n$
2. • linearity: $\mu_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$
3. • absence of multicollinearity of the $\underline{x}_j$: the covariates must be LINEARLY INDEPENDENT

Let's analyze the 3 hypotheses:

3 ABSENCE OF MULTICOLLINEARITY

What is the meaning of this hypothesis on $\underline{x}_1, \dots, \underline{x}_p$ (i.e. on the matrix X)?
Intuitively, it means that each covariate $\underline{x}_j$ should have an individual contribution for predicting Y
⇒ the information contained in $\underline{x}_j$ can NOT be derived from the other variables.

Examples of collinearity:
- the same variable is expressed using two measurement units (cm/m)
- one variable is a linear combination of the others
 (e.g. x_1 = total years of education; x_2 = years of pre-university education; x_3 = years of post-university education; $\Rightarrow x_1 = x_2 + x_3$)

What happens when this hypothesis is not satisfied?
Assume that $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$ are LINEARLY DEPENDENT:
by definition, it means that there are p scalars $a_1, \dots, a_p$, not all zero, such that

$$a_1 \underline{x}_1 + a_2 \underline{x}_2 + \dots + a_p \underline{x}_p = 0$$

This means that I can write the j-th variable as:

$$\underline{x}_j = -\frac{a_1}{a_j}\underline{x}_1 - \dots - \frac{a_{j-1}}{a_j}\underline{x}_{j-1} - \frac{a_{j+1}}{a_j}\underline{x}_{j+1} - \dots - \frac{a_p}{a_j}\underline{x}_p \qquad (*)$$

Looking now at our linear model

$$\underline{Y} = \beta_1 \underline{x}_1 + \beta_2 \underline{x}_2 + \dots + \beta_{j-1}\underline{x}_{j-1} + \beta_j \underline{x}_j + \beta_{j+1}\underline{x}_{j+1} + \dots + \beta_p \underline{x}_p + \underline{\varepsilon}$$

We can substitute $\underline{x}_j$ with $(*)$, obtaining

$$\underline{Y} = \beta_1 \underline{x}_1 + \beta_2 \underline{x}_2 + \dots + \beta_{j-1}\underline{x}_{j-1} + \beta_j\left(-\frac{a_1}{a_j}\underline{x}_1 - \dots - \frac{a_p}{a_j}\underline{x}_p\right) + \dots + \beta_p \underline{x}_p + \underline{\varepsilon}$$

$$= \underbrace{\left(\beta_1 - \beta_j \frac{a_1}{a_j}\right)}_{\beta_1^*}\underline{x}_1 + \dots + \underbrace{\left(\beta_{j-1} - \beta_j \frac{a_{j-1}}{a_j}\right)}_{\beta_{j-1}^*}\underline{x}_{j-1} + \underbrace{\left(\beta_{j+1} - \beta_j \frac{a_{j+1}}{a_j}\right)}_{\beta_j^*}\underline{x}_{j+1} + \dots + \underbrace{\left(\beta_p - \beta_j \frac{a_p}{a_j}\right)}_{\beta_{p-1}^*}\underline{x}_p + \underline{\varepsilon}$$

We have expressed the same model using only $p-1$ variables.
Hence we need to require that the covariates are linearly independent $\Rightarrow \operatorname{rank}(X) = p$

REMARK: p is the number of columns of X
If we include the INTERCEPT in the model $\Rightarrow \underline{x}_1 = \underline{1}$ constant vector of ones
Hence, to avoid collinearity, I can not have another $\underline{x}_j$ that is constant for all $i = 1, \dots, n$
⇒ we can not have $var(\underline{x}_j) = 0$ for $j = 2, \dots, p$
 This is the assumption we had in the simple lm (which included the intercept)

2 LINEARITY

$$\underline{\mu} = \sum_{j=1}^{p} \beta_j \underline{x}_j = X\underline{\beta}$$

LINEAR IN THE PARAMETERS $\underline{\beta}$

1 DISTRIBUTION: normality, homoscedasticity, incorrelation

$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$ vector of the errors

- expectation:
 $\mathbb{E}[\underline{\varepsilon}] = \underline{0}$ n-dimensional vector of zeros

- variance

$$var(\underline{\varepsilon}) = \mathbb{E}[(\underline{\varepsilon} - \cancel{\mathbb{E}[\underline{\varepsilon}]})(\underline{\varepsilon} - \cancel{\mathbb{E}[\underline{\varepsilon}]})^T] = \mathbb{E}[\underline{\varepsilon}\,\underline{\varepsilon}^T]$$

what is this quantity? $\quad \underline{\varepsilon}\,\underline{\varepsilon}^T = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix} [\varepsilon_1, \dots, \varepsilon_n] = \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & & \vdots \\ \vdots & & \ddots & \\ \varepsilon_n\varepsilon_1 & \dots & & \varepsilon_n^2 \end{bmatrix}$

$$= \begin{bmatrix} \mathbb{E}[\varepsilon_1^2] & \mathbb{E}[\varepsilon_1\varepsilon_2] & \dots & \mathbb{E}[\varepsilon_1\varepsilon_n] \\ \mathbb{E}[\varepsilon_2\varepsilon_1] & \mathbb{E}[\varepsilon_2^2] & & \vdots \\ \vdots & & \ddots & \\ \mathbb{E}[\varepsilon_n\varepsilon_1] & \dots & & \mathbb{E}[\varepsilon_n^2] \end{bmatrix}$$

since $\mathbb{E}[\varepsilon_i\varepsilon_k] = 0$ for $i \neq k$; $\mathbb{E}[\varepsilon_i^2] = \sigma^2$ for $i = 1, \dots, n$

$$= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

$(n\times n)$ matrix, diagonal elements $= \sigma^2$, off-diagonal elements $= 0$

Hence $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \quad i = 1, \dots, n \quad \Rightarrow \quad \underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

consequence for the response variable

$$\mathbb{E}[\underline{Y}] = \mathbb{E}[X\underline{\beta} + \underline{\varepsilon}] = X\underline{\beta}$$

$$var(\underline{Y}) = var(X\underline{\beta} + \underline{\varepsilon}) = var(\underline{\varepsilon}) = \sigma^2 I_n$$

Finally, the normality of $\underline{\varepsilon}$ implies the normality of $\underline{Y}$ $\Rightarrow \underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$

• INTERPRETATION OF THE COEFFICIENTS $\beta_1, \dots, \beta_p$

we have seen that in the simple linear model $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, β_2 is the expected change in Y_i (i.e., the change in $\mu_i = \mathbb{E}[Y_i]$) when we increase x_i by one unit.
(or, equivalently, the expected difference in Y when we consider two individuals i and k which differ in x by 1 unit: $\quad \beta_2 = \mathbb{E}[Y_k] - \mathbb{E}[Y_i]$, when $x_i = x_0$ and $x_k = x_0 + 1$)

How do we interpret β_j, $j = 1, \dots, p$, in the case of multiple linear regression?

$$Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

Let's consider the mean of Y of two units i and k, $\mathbb{E}[Y_i] = \mu_i$ and $\mathbb{E}[Y_k] = \mu_k$.
Assume that the values of the j-th covariate on these individuals are $x_{ij} = x_0$ and $x_{kj} = x_0 + 1$ while the other covariates are all equal: $x_{i1} = x_{k1}$, $x_{i2} = x_{k2}, \dots, x_{i,j-1} = x_{k,j-1}$, $x_{i,j+1} = x_{k,j+1}, \dots, x_{ip} = x_{kp}$

Let's now compare their means:

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip} = \beta_1 + \beta_2 x_{i2} + \dots + \beta_j x_0 + \dots + \beta_p x_{ip} \qquad \text{mean of individual } i$$

$$\mu_k = \beta_1 + \beta_2 x_{k2} + \dots + \beta_j x_{kj} + \dots + \beta_p x_{kp} = \beta_1 + \beta_2 x_{k2} + \dots + \beta_j (x_0 + 1) + \dots + \beta_p x_{kp} = \beta_1 + \beta_2 x_{k2} + \dots + \beta_j x_0 + \beta_j + \dots + \beta_p x_{kp} \qquad \text{mean of individual } k$$

If we study the difference in their means

$$\Rightarrow \mu_k - \mu_i = \beta_j$$

β_j now represents the expected change in Y_i (i.e., the change in μ_i), when we increase x_{ij} by one unit, while keeping all other covariates fixed.