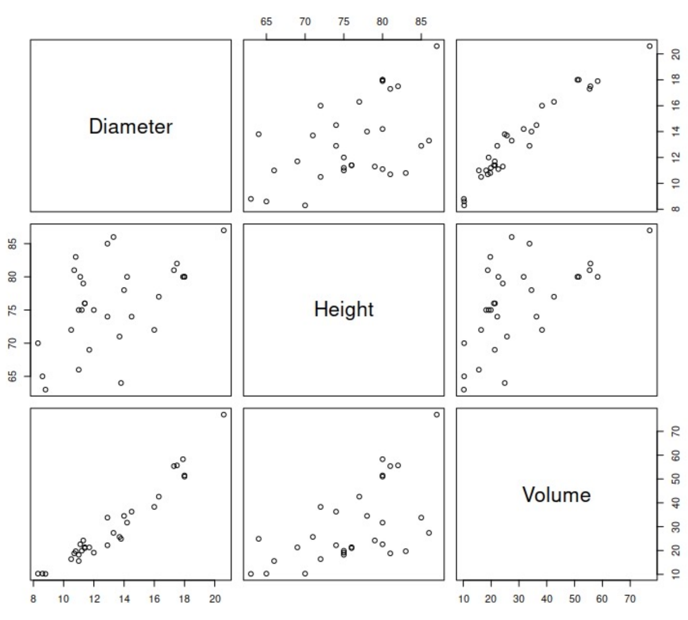
MULTIPLE LINEAR REGRESSION

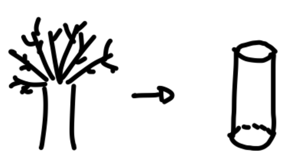
There are now $p>1$ covariates $x_1, \dots, x_p$.

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

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



Actually, if we think of the shape of a tree, we could think of approximating it to a cylinder

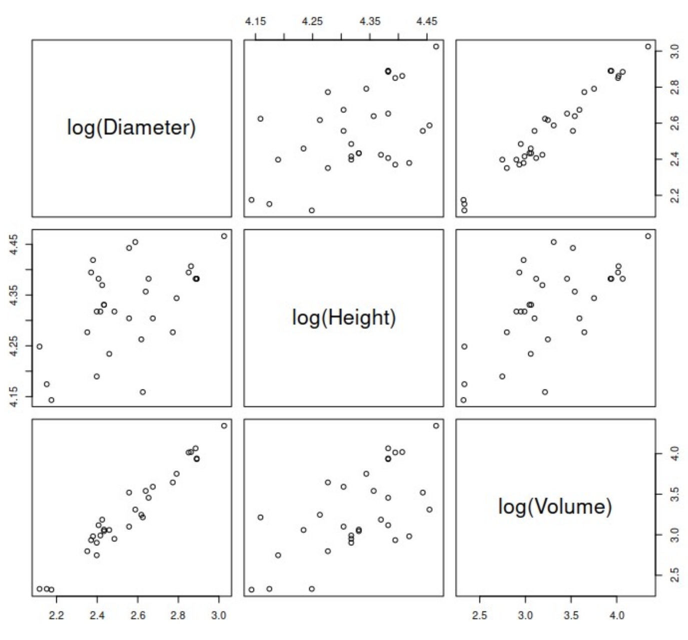


$$\text{volume} = \pi \cdot \text{radius}^2 \cdot \text{height}$$
$$= \pi \cdot (d/2)^2 \cdot \text{height} \quad \text{(not linear!)}$$

but

$$\log(\text{volume}) = \log \pi + 2\log d - 2\log 4 + \log \text{height}$$

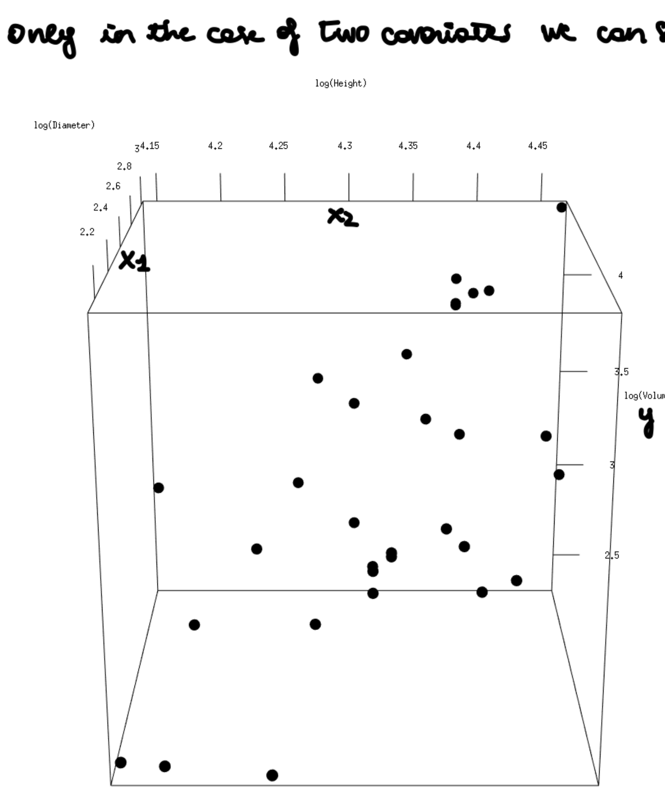
We can consider the transformed variables

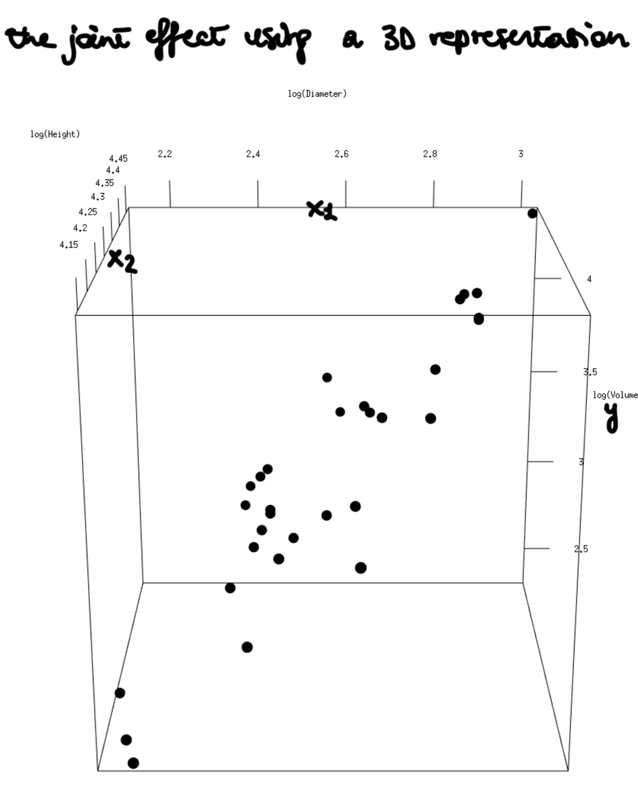


$Y = \log(\text{volume})$
$x_1 = \log(\text{diameter})$
$x_2 = \log(\text{height})$

with 2 or more covariates we can no longer see the JOINT effect they have on y, but only the INDIVIDUAL effect of 1 predictor if we use a scatterplot.
The goal of the multiple lm is to study the JOINT EFFECT of the covariates on y.

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





MODEL SPECIFICATION

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

$$Y_i = \mu_i + \varepsilon_i$$
$$= \beta_1 \underset{\substack{\| \\ 1 \ \forall i \text{ if we include the intercept}}}{x_{i1}} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1, \dots, n$$

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

1. • normality, homoscedasticity, corr = 0 $\rightarrow$ $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ $\quad i = 1, \dots, n$
2. • linearity: $\mu_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$
3. • absence of multicollinearity of the x_j: the covariates must be linearly independent
   (in the simple lm we had an analogous assumption: $\text{var}(x) \neq 0$)

notation:

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

for $j = 1, \dots, p$

$$\Rightarrow \underline{Y} = \beta_1 \underline{x}_1 + \dots + \beta_p \underline{x}_p + \underline{\varepsilon}$$
$$\Rightarrow \underline{Y} = \sum_{j=1}^{p} \beta_j \underline{x}_j + \underline{\varepsilon}$$
$$\Rightarrow \underset{n\times 1}{\underline{Y}} = \underbrace{\underset{n\times p}{X} \underset{p\times 1}{\underline{\beta}}}_{\underline{\mu} \ n\times 1} + \underset{n\times 1}{\underline{\varepsilon}}$$

with

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

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

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

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

$\underline{Y}$ is a vector of r.v.
X is a matrix of constants (known)
$\underline{\beta}$ is a vector of constants (unknown)
$\underline{\varepsilon}$ is a vector of r.v.

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
$\Rightarrow$ 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 $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$ are linearly dependent: this 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 $\underline{x}_j = -\frac{a_1}{a_j}\underline{x}_1 - \dots - \frac{a_{j-1}}{a_j}\underline{x}_{j-1} - \frac{a_{j+1}}{a_j}\underline{x}_{j+1} - \dots - \frac{a_p}{a_j}\underline{x}_p$

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

We have expressed the same model using only $p-1$ variables.
Hence we need to require that the covariates are linearly independent $\Rightarrow$ $\text{rank}(X) = p$ (rank = # columns including the intercept! $\underline{x}_1 = \underline{1}$)

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

• DISTRIBUTION: normality, homoscedasticity, incorrelation

$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$ $\quad \mathbb{E}[\underline{\varepsilon}] = \underline{0}$ n-dimensional vector of zeros

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

$$\underset{n\times n}{\text{var}(\underline{\varepsilon})} = \mathbb{E}[(\underline{\varepsilon} - \mathbb{E}[\underline{\varepsilon}])(\underline{\varepsilon} - \mathbb{E}[\underline{\varepsilon}])^T]$$
$$= \mathbb{E}[\underline{\varepsilon}\underline{\varepsilon}^T] \ *$$
$$= \sigma^2 I_n$$
$$= \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots & \sigma^2 \end{bmatrix}$$

$$* \ \underline{\varepsilon}\underline{\varepsilon}^T = \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & & \vdots \\ \vdots & & \ddots & \\ \varepsilon_n\varepsilon_1 & \dots & & \varepsilon_n^2 \end{bmatrix}$$

since $\mathbb{E}[\varepsilon_i \varepsilon_k] = 0$
$\mathbb{E}[\varepsilon_i^2] = \sigma^2$

$$\Rightarrow \text{var}(\underline{Y}) = \sigma^2 I_n$$

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

$$Y = \beta_1 + \beta_2 x + \varepsilon$$

β_2 is the change in μ when we change x of one unit.
How do we interpret β_j, $j = 1, \dots, p$, in the case of multiple linear regression?

$$Y = \beta_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

β_j is now the change in μ when we change x_j of one unit, while keeping all other covariates fixed.
Let's consider the mean μ at two points x_j and $(x_j + 1)$

$$\mu^{(1)} = \beta_1 + \beta_2 x_2 + \dots + \beta_j x_j + \dots + \beta_p x_p$$
$$\mu^{(2)} = \beta_1 + \beta_2 x_2 + \dots + \beta_j (x_j + 1) + \dots + \beta_p x_p$$
$$\Rightarrow \mu^{(2)} - \mu^{(1)} = \beta_j$$