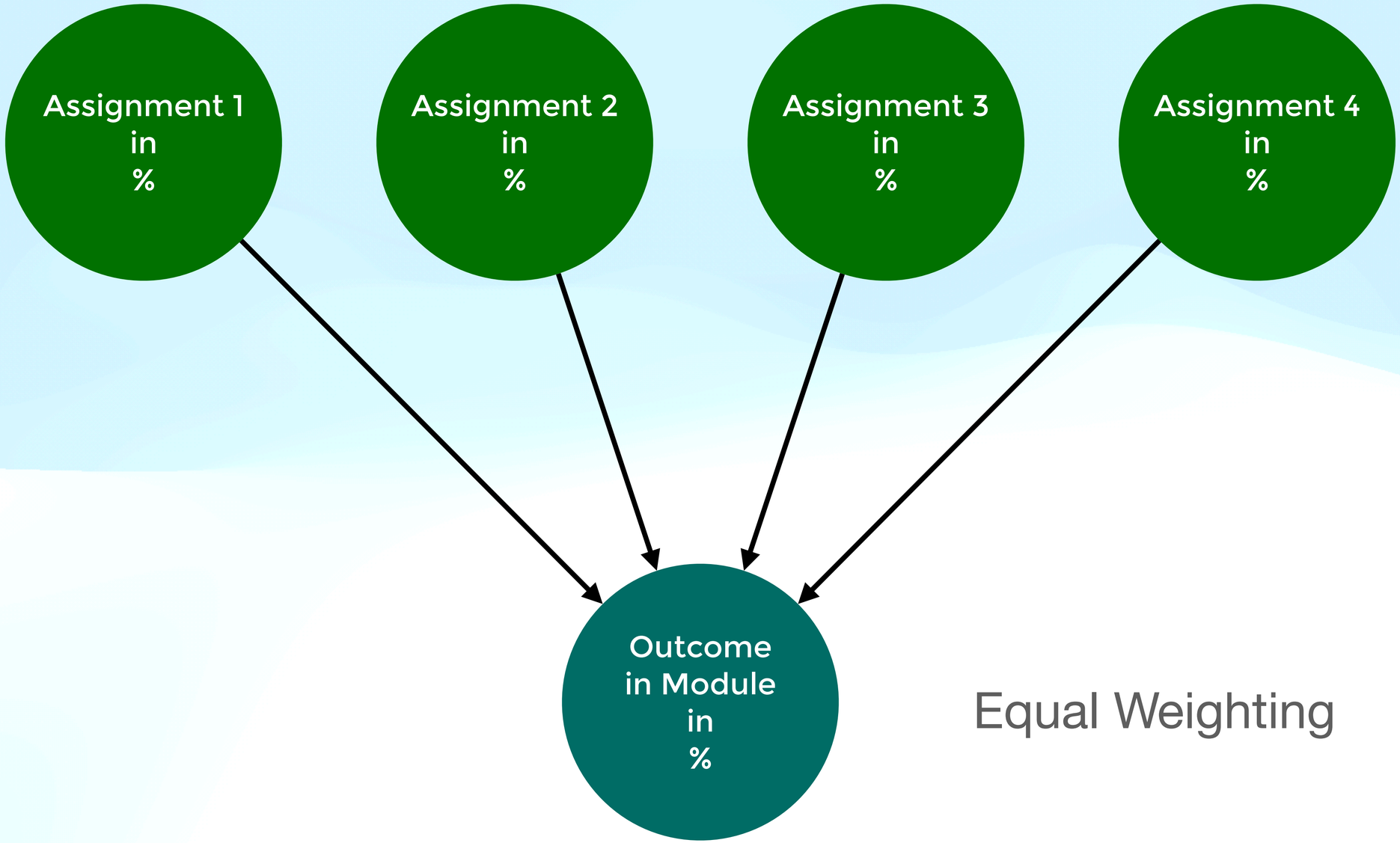
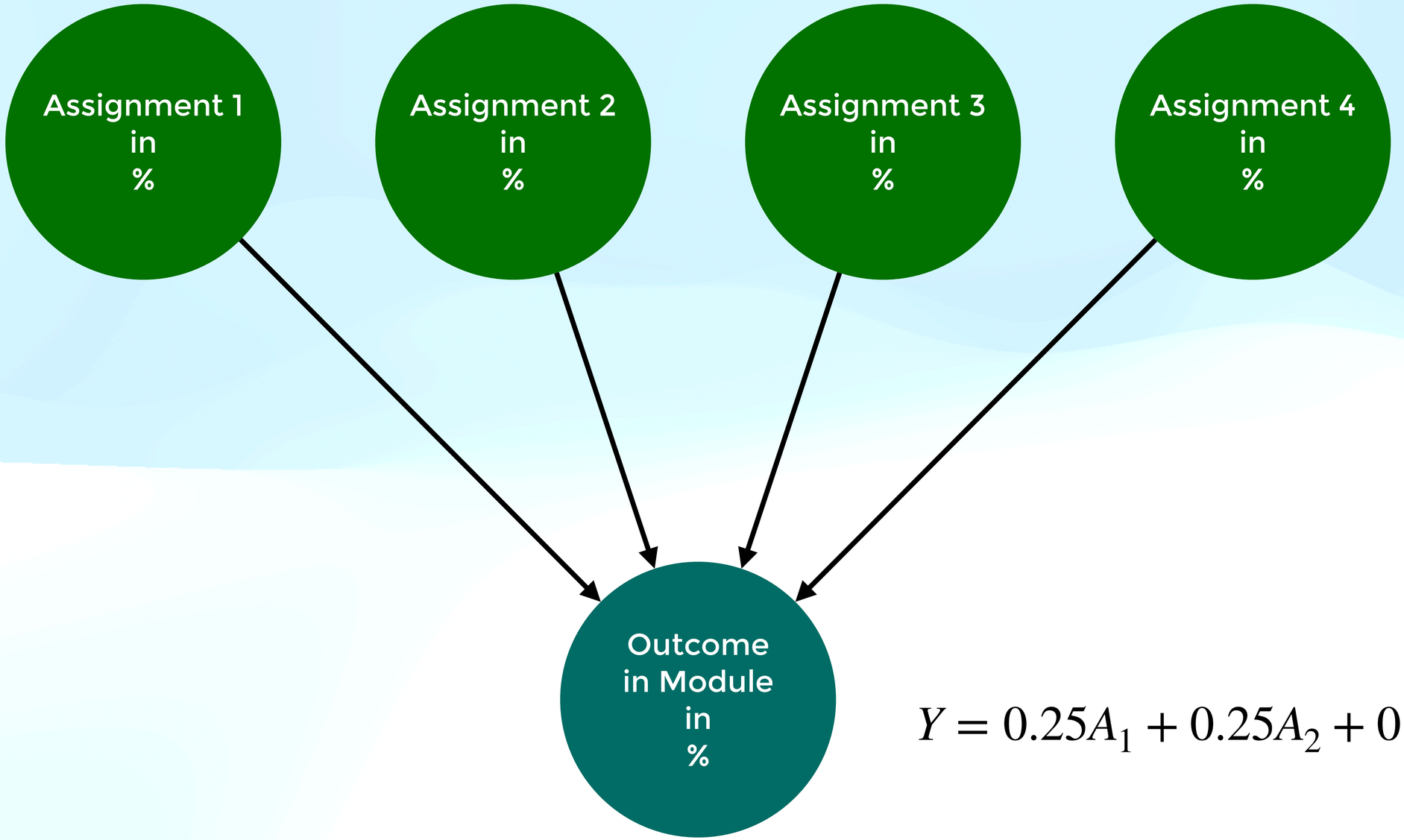
Regression and Simulation Methods

Week 2: Linear Models

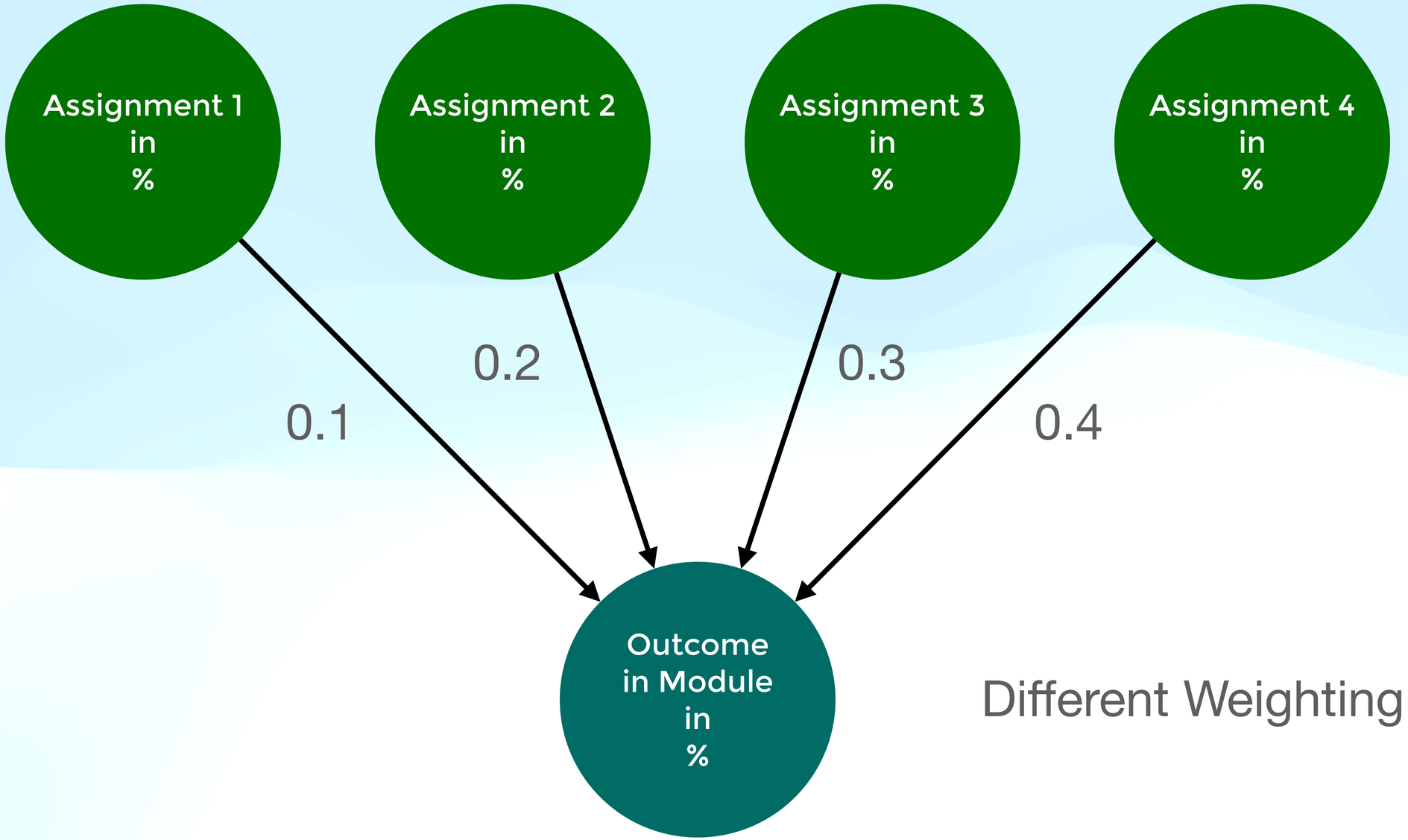
Chris Oldnall, 17th October 2023



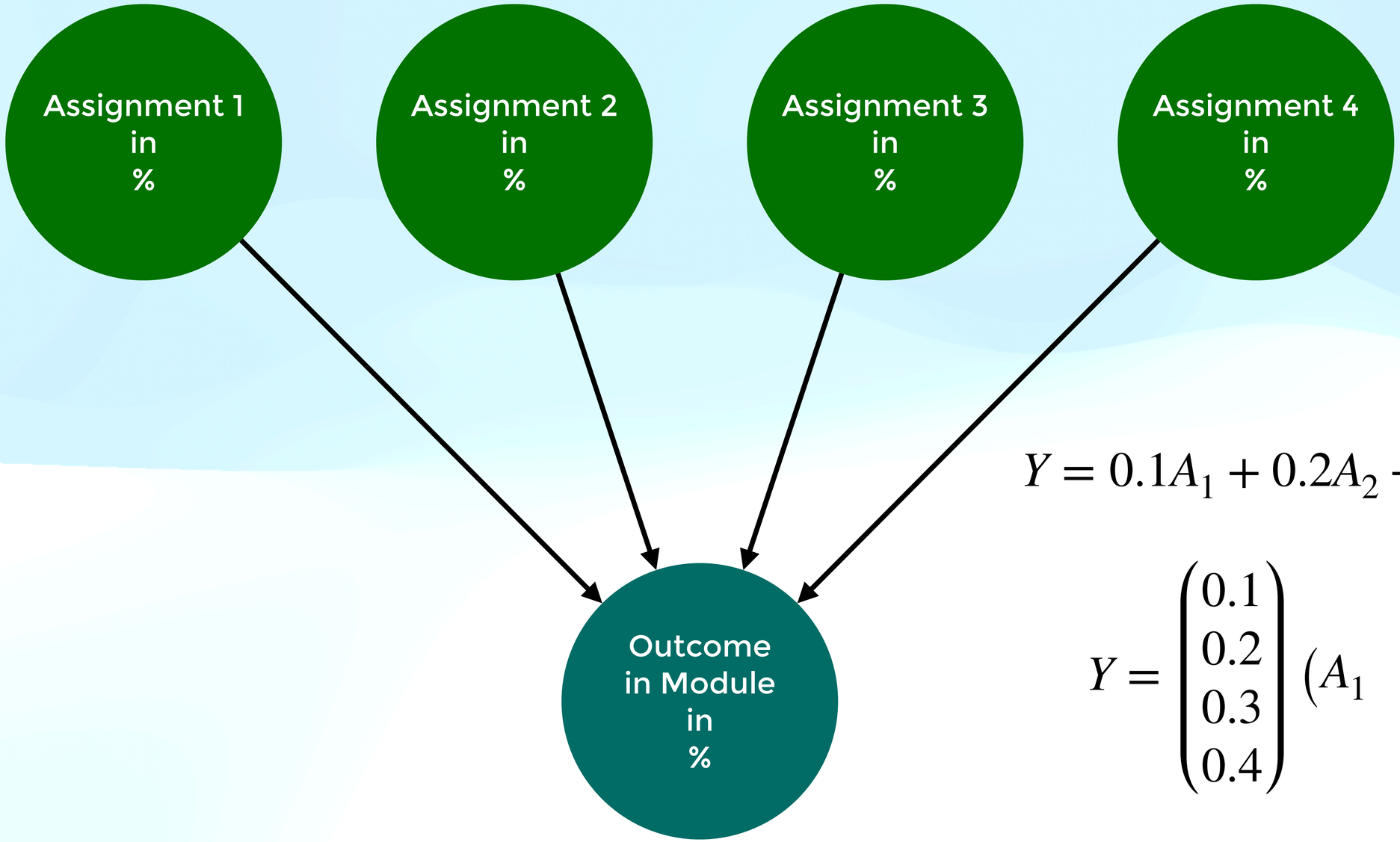
Equal Weighting



$$Y = 0.25A_1 + 0.25A_2 + 0.25A_3 + 0.25A_4$$



Different Weighting



$$Y = 0.1A_1 + 0.2A_2 + 0.3A_3 + 0.4A_4$$

$$Y = \begin{pmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{pmatrix} \begin{pmatrix} A_1 & A_2 & A_3 & A_4 \end{pmatrix}$$

$$Y = \beta \mathrm{A}$$

4 Steps to (Normal) Linear Modelling

- Formation
- Estimation
- Checking
- Analysis

Formation

Explanatory Variables (#p-1)

Error/natural variation

Outcome (#1)

$$y = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_{p-1} x_{p-1,i} + \epsilon_i$$

Parameters (#p)

Amount of change in the mean value of y when the x_j explanatory variable increases by one unit and the other explanatory variables are held fixed.

Formation Assumptions

i. The relationship between y and x_j is linear if all other explanatory variables are held fixed.

ii. We assume for the error that,

$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2)$$

Notational Convenience

$$Y = X\theta + \epsilon$$

$$\overbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}^{\text{Outcome}} = \overbrace{\begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{p-1,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{p-1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \cdots & x_{p-1,n} \end{pmatrix}}^{\text{Design Matrix (n x p)}} \overbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}}^{\text{Parameter Vector}} + \overbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}}^{\text{Errors}}$$

Estimation of Parameters

Typically in this text, estimation is referred to as ‘fitting’. This is a language preference - however I will always refer to estimation.

The most common procedure for estimation of parameters in a linear model is ordinary least squares (OLS). This is the process of solving the quadratic minimisation problem.

$$
\begin{aligned}
L(\mathcal{O}, \theta) &= \|Y - X\theta\|^2 \\
\hat{\theta} &= \arg\min_{\theta} L(\mathcal{O}, \theta) \\
&= \arg\min_{\theta} \|Y - X\theta\|^2
\end{aligned}
$$

For further justification, see ‘Data Science for Dynamical Systems’ SMSTC course by S.Klaus in 2023

Estimation of Parameters

$$
\begin{aligned}
L(\mathcal{O}, \theta) &= \|Y - X\theta\|^2 \\
&= (Y - X\theta)^T (Y - X\theta) \\
&= Y^T Y - Y^T X\theta - \theta^T X^T Y + \theta^T X^T X\theta \\
\frac{\partial L(\mathcal{O}, \theta)}{\partial \theta} &= -2X^T Y + 2X^T X\theta \\
0 &= -2X^T Y + 2X^T X\hat{\theta} \\
X^T Y &= X^T X\hat{\theta} \\
\hat{\theta} &= (X^T X)^{-1} X^T Y
\end{aligned}
$$

Least Squares Properties

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}[(X^T X)^{-1} X^T Y]$$
$$= (X^T X)^{-1} X^T \mathbb{E}[Y]$$
$$= (X^T X)^{-1} X^T \mathbb{E}[X\theta + \epsilon]$$
$$= (X^T X)^{-1} X^T X\theta$$
$$= \theta$$

$$\mathbb{V}[\hat{\theta}] = ?$$

Unbiased estimate.

Least Squares Properties

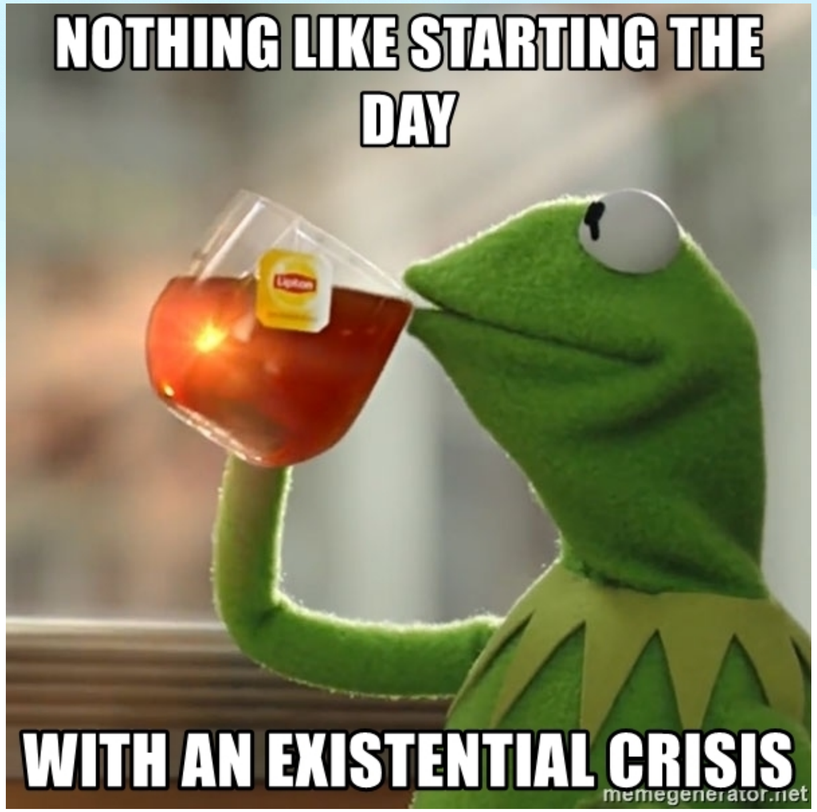
$$
\begin{aligned}
\mathbb{E}[\hat{\theta}] &= \mathbb{E}[(X^T X)^{-1} X^T Y] \\
&= (X^T X)^{-1} X^T \mathbb{E}[Y] \\
&= (X^T X)^{-1} X^T \mathbb{E}[X\theta + \epsilon] \\
&= (X^T X)^{-1} X^T X\theta \\
&= \theta
\end{aligned}
$$

Unbiased estimate.

$$
\begin{aligned}
\mathbb{V}[\hat{\theta}] &= \mathbb{V}[(X^T X)^{-1} X^T Y] \\
&= (X^T X)^{-1} X^T \mathbb{V}[Y] X (X^T X)^{-1} \\
&= (X^T X)^{-1} X^T \mathbb{V}[X\theta + \epsilon] X (X^T X)^{-1} \\
&= (X^T X)^{-1} X^T \mathbb{V}[\epsilon] X (X^T X)^{-1} \\
&= (X^T X)^{-1} X^T \sigma^2 X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

Checking

How do we know that this was the correct model to use for our data?



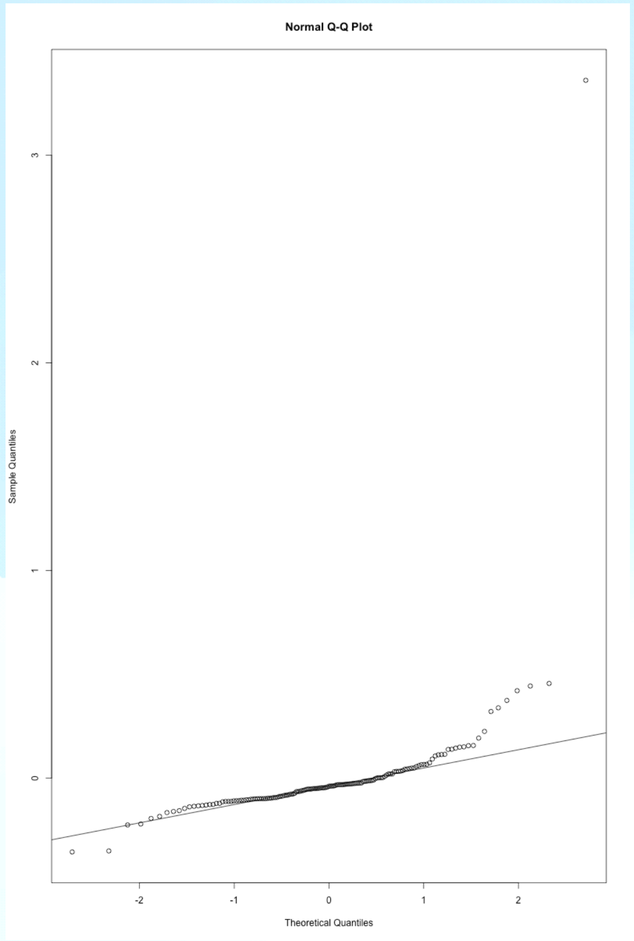
Checking

We can use the residuals (the difference between the ‘true’ and fitted model) to assess the model fit. This is just a linear transformation of the errors.

$$\begin{aligned}\hat{\epsilon} &= y - \hat{y} \\ &= (I - H)y \\ &= (I - H)(X\beta + \epsilon) \\ &= (I - H)\epsilon\end{aligned}$$

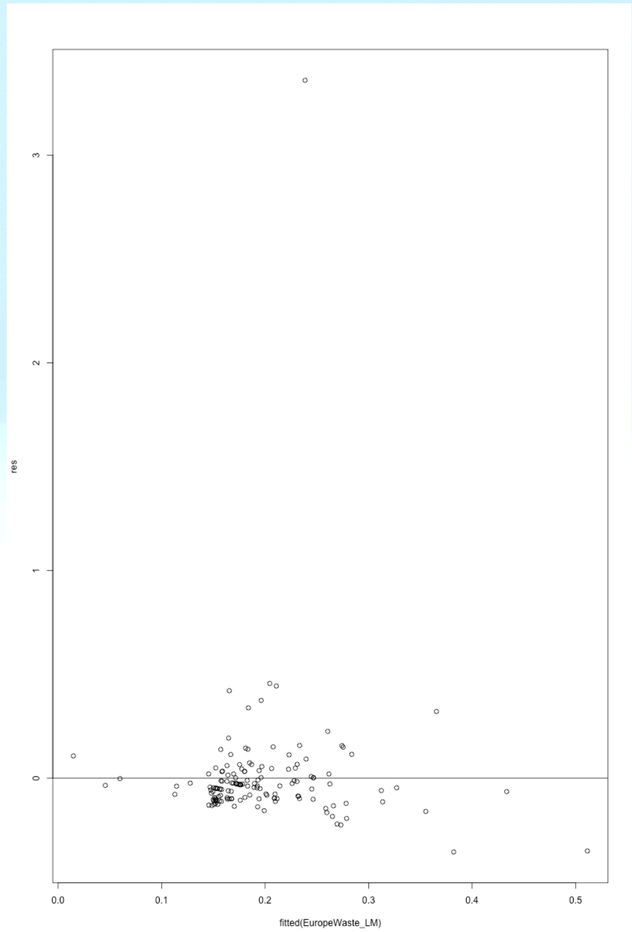
It is expected that these quantities would follow a normal distribution. Therefore QQ-plots are used to assess if the residuals look ‘ok’.

An example of a normal QQ-plot.



Looking for a straight line ✅

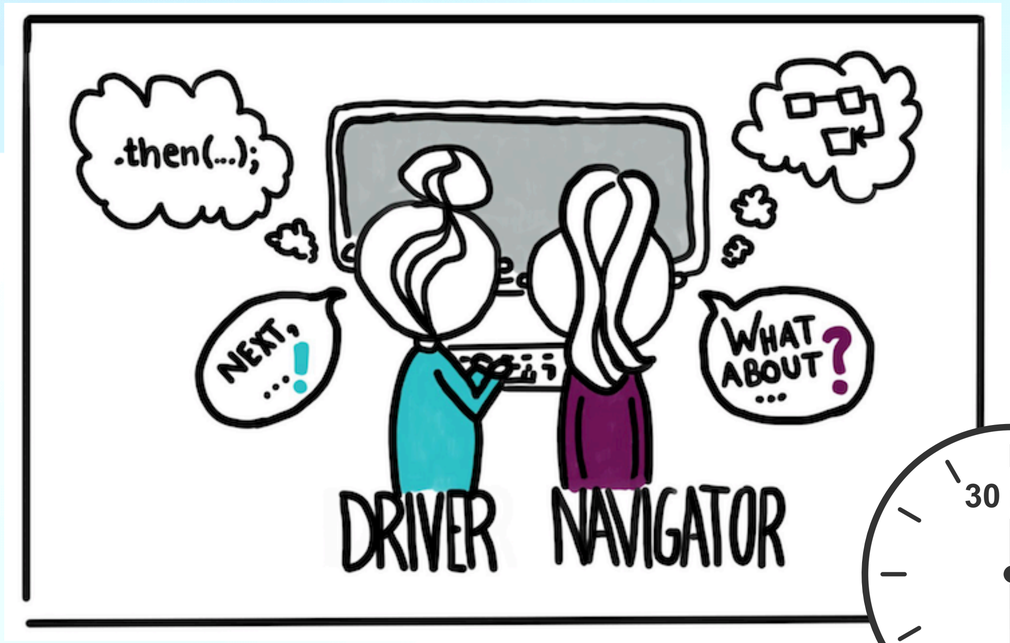
Residuals vs fitted values.

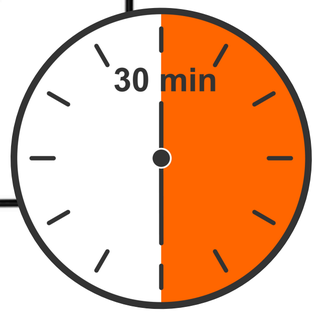


Looking for complete random distribution ✅

Rest of the tutorial...

- In pairs work on the second notebook. Paired programming will continue!





Rest of the week...

Read ahead in 2.6 to see what the definition of RSS.
Then answer

2-1: Prove that $\mathbb{E}[\text{RSS}] = (n-p)\sigma^2$

2-3:
A useful graphical method for assessing the potential value of adding a new variable to a linear model is an added variable plot. Consider the DO data (found on the SMSTC website) for Station 20, with explanatory variables Temperature and Salinity.

Construct the residuals for this model. These residuals could be plotted against the new variables; try this with Year. However, some theoretical analysis shows that a better thing to do is to first find the residuals of the linear model which has Year as response and Temperature, Salinity as explanatory variables. This is simply a device to extract the information on Temperature and Salinity from Year, before we assess its further value.

The added variable plot is then constructed by plotting the residuals of the DO model against the residuals for the Year model. Try this, and assess the degree to which Year has further useful information. Can you give an intuitive explanation of why the added variable plot may be better than simply plotting the residuals of the first model against Year?