Exercise: Dummy encoding and linear models

(Source: (**?**, p151).)

Consider a linear regression model of the form

$$y_i = w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + w_4 x_{i4} + \epsilon_i \quad (1)$$

where $x_{i1} = 1$, $x_{i2} = g_i \in \{0, 1\}$ specifies if person i is in a control group or not, $x_{i3} =$ age of person i, and $x_{i4} = x_{i2} \times x_{i3}$. So we have

$$\mathbb{E}[y|\mathbf{x}_i, g_i = 0] = w_1 + w_3 \text{age}_i \quad (2)$$

$$\mathbb{E}[y|\mathbf{x}_i, g_i = 1] = (w_1 + w_2) + (w_3 + w_4)\text{age}_i \quad (3)$$

Hence the difference in offsets between the two groups is w_2, and the difference in slopes is w_4. Sketch the regression line for the two groups assuming $w_1 = 1$, $w_3 = 1$ and with the following settings for the other parameters: (1) $w_2 = 0$, $w_4 = 0$, (2) $w_2 = 0$, $w_4 = 1$, (3) $w_2 = 1$, $w_4 = 0$, (4) $w_3 = 1$, $w_4 = 1$. You should have 4 figures, each with 2 lines. You can draw the figures by hand, or use Matlab. Assume the age ranges from 0 to 10.