1 General Notes

There is a nice example in Gelman page 53 on how different units affect coefficient interpretation. Compare the following

$$earnings = -61000 + 51 \cdot height[inmillimetres] + \epsilon \tag{1}$$

And

$$earnings = -61000 + 81000000 \cdot height[inmiles] + \epsilon \tag{2}$$

It has been given that standard deviation of height is equal to 3.8 inches, which is 97 millimetres or 0.000061 miles. Observe that we obtain the same expected difference in earnings for the matching units

$$51 \cdot 97 = 81000000 \cdot 0.000061 = 4900 \tag{3}$$

2 Scaling

Suppose there is this model

$$\log_e arn = height + male \tag{4}$$

Once we fit a linear regression model we have received the following coefficients:

$$\log_e arn = 8.153 + 0.021 height + 0.423 male \tag{5}$$

Also,

$$\exp(0.021) = 1.02 \tag{6}$$

Therefore, for two people of the same gender one inch of height contributes towards 2% of salary increase. Say that the standard errors are as follows:

intercept	height	male
0.603	0.009	0.072

A 70-inch tall person will have earnings equal to $8.153 + 0.021 \cdot 70 = 9.623$. Knowing that predictive standard deviation is 0.88, there is 68 chance the person earning are within $9.623 + / - 0.88 = [8.74, 10.50] = [\exp(8.74), \exp(10.50)] = [6000, 36000]$. The R^2 value for this model was pretty low therefore it is expected to observe such a wide salary range.

3 Interaction

Consider the following model

$$\log(earnings)\ (height) \tag{7}$$

Once we have fit the equation coefficients given data we obtain the following:

1. intercept - 5.74

2. height - 0.06

The difference of 0.06 in height corresponds to $\exp(0.06) = 1.062$. Therefore, a unite change of β_1 corresponds to a 6% increase in the y value. The opposite holds for negative values (i.e there is 6% decrease in the outcome value for a drop of one unit in β_1). Observe that It would be interesting to see if gender distinction contributes to the 6% increase.

Suppose there is this equation

$$\log(e) = 8.4 + 0.017 \cdot h - 0.079 \cdot m + 0.007 \cdot h \cdot m \tag{8}$$

where e corresponds to earnings, h height, m to male and $h \cdot m$ it an interaction term between height and male. Now, observe that male intercept does not really have a proper interpretation as it expect the height being equal to 0. However, the interaction term corresponds to the " difference in the slopes for log earnings between men and women"