Interpreting the slope and intercept in a linear regression model

Example 1. Data were collected on the depth of a dive of penguins and the duration of the dive. The following linear model is a fairly good summary of the data, where *t* is the duration of the dive in minutes and *d* is the depth of the dive in yards. The equation for the model is $d = 0.015 + 2.915t$

Interpret the slope: If the duration of the dive increases by 1 minute, we predict the depth of the dive will increase by approximately 2.915 yards.

Interpret the intercept. If the duration of the dive is 0 seconds, then we predict the depth of the dive is 0.015 yards.

Comments: The interpretation of the intercept doesn't make sense in the real world. It isn't reasonable for the duration of a dive to be near $t = 0$, because that's too short for a dive. If data with x-values near zero wouldn't make sense, then usually the interpretation of the intercept won't seem realistic in the real world. It is, however, acceptable (even required) to interpret this as a coefficient in the model.

Example 2. Reinforced concrete buildings have steel frames. One of the main factors affecting the durability of these buildings is carbonation of the concrete (caused by a chemical reaction that changes the pH of the concrete) which then corrodes the steel reinforcing the building. Data is collected on specimens of the core taken from such buildings, where the depth of the carbonation, in mm, called d, and the strength of the concrete, in Mpa, called *s*, are measured. It is found that the model is $s = 24.5 - 2.8 \cdot d$

Interpretation of the slope: If the depth of the carbonation increases by 1 mm, then the model predicts that the strength of the concrete will decrease by approximately 2.8 Mpa.

Interpretation of the intercept: If the depth of the carbonation is 0, then the model predicts that the strength of the concrete is approximately 24.5 Mpa.

Comments: Notice that it isn't necessary to fully understand the units in which the variables are measured in order to correctly interpret these coefficients. While it is good to understand data thoroughly, it is also important to understand the structure of linear models.

In this model, notice that the strength decreases as the carbonation increases, which is shown by the negative slope coefficient. When you interpret a negative slope, notice that you must say that, as the explanatory variable increases, then the response variable decreases.

Example 3. When cigarettes are burned, one by-product in the smoke is carbon monoxide. Data is collected to determine whether the carbon monoxide emission can be predicted by the nicotine level of the cigarette. It is determined that the relationship is approximately linear when we predict carbon monoxide, *C*, from the nicotine level, *N*. Both variables are measured in milligrams. The formula for the model is
$C = 3.0 + 10.3 \cdot N$

Interpret the slope: If the amount of nicotine goes up by 1 mg, then we predict the amount of carbon monoxide in the smoke will increase by 10.3 mg.

Interpret the intercept: If the amount of nicotine is zero, then we predict that the amount of carbon monoxide in the smoke will be about 3.0 mg.

Example 4. Data was collected to determine the length a golf ball will travel when hit by a golf club with a certain speed. The speed, s, is measured in miles per hour and the length the ball travels, d, is measured in yards. The following formula gives the relationship $d = 3.18 + 57.66 \cdot s$

Interpret the slope: If the speed of the club hitting the ball increases by 1 mph, then the model predicts that the length the ball travels increases by 57.66 yards.

Interpret the intercept: If the ball is hit with a speed of 0 mph, then the model predicts that the length the ball travels will be 3.18 yards.

Comments: The interpretation of the intercept doesn't make sense because speed of 0 mph means that the ball wasn't hit at all. Of course, 3.18 yards isn't very far either, so it looks quite unrealistic as well. There would have been no data with speed near 0 mph because for this to be a reasonable hit for a golf ball, the speed would have to be considerably larger than 0 mph. This is a correct interpretation of the coefficient in the model, even though the intercept is not a reasonable data point.