# Lesson Learning Objectives

**LO 1.** Define a leverage point as a point that lies away from the center of the data in the horizontal direction.

**LO 2.** Define an influential point as a point that influences (changes) the slope of the regression line.

- This is usually a leverage point that is away from the trajectory of the rest of the data.

**LO 3.** Do not remove outliers from an analysis without good reason.

**LO 4.** Be cautious about using a categorical explanatory variable when one of the levels has very few observations, as these may act as influential points.

**LO 5.** Determine whether an explanatory variable is a significant predictor for the response variable using the t-test and the associated p-value in the regression output.

**LO 6.** Set the null hypothesis testing for the significance of the predictor as $H_0 : \beta_1 = 0$, and recognize that the standard software output yields the p-value for the two-sided alternative hypothesis.

- Note that $\beta_1 = 0$ means the regression line is horizontal, hence suggesting that there is no relationship between the explanatory and response variables.

**LO 7.** Calculate the T score for the hypothesis test as

$$T_{df} = \frac{b_1 - \text{null value}}{SE_{b_1}}$$

with $df = n - 2$.

- Note that the T score has $n - 2$ degrees of freedom since we lose one degree of freedom for each parameter we estimate, and in this case we estimate the intercept and the slope.

**LO 8.** Note that a hypothesis test for the intercept is often irrelevant since it's usually out of the range of the data, and hence it is usually an extrapolation.

**LO 9.** Calculate a confidence interval for the slope as

$$b_1 \pm t^\star_{df} SE_{b_1},$$

where $df = n - 2$ and $t^\star$ is the critical score associated with the given confidence level at the