Questions Module 5.4

Suppose a categorical response variable, **Defective**, can have only two possible values: *Defective* and *Nondefective*.

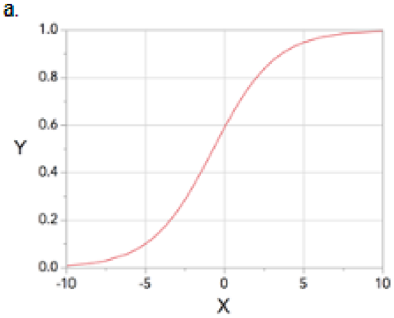
True or False: If the probability of *Defective* is 0.20, then the probability of *Nondefective* must be 0.80.

- *a.* True
- *b.* False
- *c.* There is not enough information to answer this question.

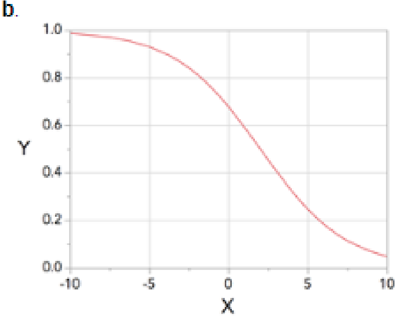
Incorrect.
The correct answer is **a**. The statement is true because the sum of the probabilities of all possible outcomes is 1.0. The probability of each individual outcome, *Defective* or *Nondefective*, can be between 0 and 1.

Run the **Logistic.jsl** script. This script enables you to change the values of the slope and intercept to explore the shape of the logistic function. Use this script to answer the matching question below. Match each description below to the appropriate logistic curve.

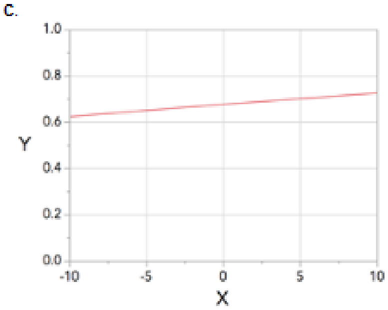
b negative slope	a.



c slope close to zero	b.



a positive slope	c.



Correct.

a. The logistic curve has an upward slope, which means that ***higher*** X values are associated with a ***higher*** probability of the event happening.

b. The logistic curve has a downward slope, which means that ***higher*** X values are associated with a ***lower*** probability of the event happening.

c. The slope of the logistic curve is relatively ***flat***, which means that there is not much change in the probability of the event happening for any X values.

You fit a logistic regression model for **Outcome** as a function of the five predictors. Pretend, for this example, that the raw **Impurity** measurements are not available.

Temp, **Catalyst Conc**, **Reaction Time**, and **Reacto**r are all significant. If the ultimate goal is to eliminate batches that failed to meet specifications, how can you use these results to achieve the goal?

There are many possible answers. You can use the Prediction Profiler to find settings of **Temp**, **Catalyst Conc**, and **Reaction Time** in which the predicted probability of fail is 0. There might be different settings that enable you to achieve this goal.

We might fit a model including interactions, to determine whether you can learn more about potential causes of failures.

We might want to design an experiment to find and confirm optimal settings for these factors.

If available, you might want to include other information, such as cost and cycle time. You might be able to find settings that minimize failures and also reduce cost and cycle time.

There is also a difference among the reactors. **Reactor 3** has the highest probability of **Fail**. You might want to investigate these differences to determine the root causes. It might be possible to find the settings of the continuous predictors that are robust, or insensitive, to the differences among the reactors.

For the Metal Parts scenario, recall that defective parts are parts with uneven coating. You are on a project team that is tasked with reducing the defective rate.

Use logistic regression (among other problem-solving tools) to identify the potential causes of defective parts.

With logistic regression, you identified six important factors and three two-way interactions.

Effect Summary

Source	LogWorth	PValue
nozzle size*paint supplier	6.294	0.00000
paint viscosity (cP)*belt speed (ft/sec)	5.370	0.00000
belt speed (ft/sec)	3.611	0.00024 ^
nozzle size	2.492	0.00322 ^
paint supplier	2.005	0.00988 ^
pump pressure (psi)	1.660	0.02190
paint viscosity (cP)	1.569	0.02698 ^
humidity	1.354	0.04427
pump pressure (psi)*paint supplier	1.208	0.06197

However, the misclassification rate for your logistic model is 23.5%, and the model does a poor job of classifying defective batches. Only 69 of the 127 defective batches, or 54%, were found by the model.

Confusion Matrix

Training

Actual Outcome	Predicted Count Defective	Good
Defective	69	58
Good	32	224

Although you found important variables, there is still much that you don't know about what causes uneven coating.

What are some potential next steps?

One possible approach is to take a step back and apply the methods that were introduced in the Statistical Thinking and Problem Solving module. For example, if you haven't already done so, develop a process map to better understand the coating process, and use a cause-and-effect diagram to brainstorm the potential causes of uneven coating. Collect new data about the potentially important variables that were overlooked in the initial analysis.

Was the measurement system evaluated? Are the measurement systems for the response and the predictors capable? Remember that, in the White Polymer case study, you initially couldn't solve the problem with the existing data because the measurement systems were bad. If the measurement systems are not capable, then the data might not be useful for solving the problem. For a discussion of measurement systems analysis, see the Quality Methods module.

Is there a continuous measure of coating uniformity that can be used instead of classifying the parts as *good* or *defective*? Continuous responses provide more information content than categorical responses. With continuous data, you can estimate measures of the centering and spread of the response. That enables you to more easily detect small shifts or changes in the process.

Another option is to conduct an experiment to study the important effects that were identified and determine optimal settings for these variables. You'll learn about designed experiments in the next module.