**Simple Explanation of Logistic Regression**

**Answer 1 -** The output for linear regression is a number that has its real meaning.  
  
The output for a logistic regression is a number that represents the probability of the event happening (i.e. the probability of people clicking an ad online, the probability of death in titanic disaster, etc.)  
  
The intuition behind logistic regression is to transform the output of a linear regression which has a wider range, to a range that probability lies in, which is [0,1]. The transformation formula is [Logit](https://en.wikipedia.org/wiki/Logit) that maps a value to a number in the range [0,1].

**Answer 2 –**

The logistic regression model is maximum likelihood using the natural parameter (the log-odds ratio) to contrast the relative changes in the risk of the outcome per unit difference in the predictor. This is assuming, of course, a binomial probability model for the outcome. That means that the consistency and robustness properties of logistic regression extend directly from maximum likelihood: robust to missing at random data, root-n consistency, and existence and uniqueness of solutions to estimating equations. This is assuming the solutions are not on the boundaries of parameter space (where log odds ratios are ±∞±∞

). Because logistic regression is maximum likelihood, the loss function is related to the likelihood, since they're equivalent optimization problems.

With quasilikelihood or estimating equations (semiparametric inference), existence, uniqueness properties still hold but the assumption that the mean model holds is not relevant and the inference and standard errors are consistent regardless of model misspecification. So in this case, it's not a matter of whether the sigmoid is the correct function, but one that gives us a trend that we can believe in and is parameterized by parameters that have an extensible interpretation.

The sigmoid, however, is not the only such binary modeling function around. The most commonly contrasted probit function has similar properties. It doesn't estimate log-odds ratios, but functionally they look very similar and tend to give very similar approximations to the *exact same thing*. One need not use boundness properties in the mean model function either. Simply using a log curve with a binomial variance function gives relative risk regression, an identity link with binomial variance gives additive risk models. All this is determined by the user. The popularity of logistic regression is, sadly, why it's so commonly used. However, I have my reasons (the ones that I stated) why I think it's well justified for it's use in most binary outcome modeling circumstances.

In the inference world, for rare outcomes, the odds ratio can be roughly interpreted as a "relative risk", i.e. a "percent relative change in the risk of outcome comparing X+1 to X". This isn't always the case and, in general, an odds ratio cannot and should not be interpreted as such. However, that parameters have interpretation and can be easily communicated to other researchers is an important point, something sadly missing from the machine learnists' didactic materials.

The logistic regression model also provides the conceptual foundations for more sophisticated approaches such as hierarchical modeling, as well as mixed modelling and conditional likelihood approaches which are consistent and robust to exponentially growing numbers of nuisance parameters. GLMMs and conditional logistic regression are very important concepts in high dimensional statistics.

Source: [Intuition behind logistic regression](http://stats.stackexchange.com/questions/71176/intuition-behind-logistic-regression)