1. Allows you to assess how generalizable your model is (by testing it on data it hasn't seen yet) and avoid overfitting
2. First model -- high variance, Second model -- high bias
3. High variance
4. False
5. When the data set is very small -- not enough data to split into training/validation/test
6. False
7. Do not use in analysis (drop the variable)
8. Substituting a value in place of missing value (typically a guess of the true value). You can replace missing values with median/mode/mean
9. Equal width: same range of values per bin (eg: bin 1: 1- 10, bin 2: 11-20)
   Equal depth: percentiles. Same number of observations per bin.
10. For changes of <=20%, there's an approximately one-to-one relationship. So the translation will match that of a regression without logs.

    This is not a correct interpretation of what is going on – But I want to leave it here so that I can point out the following correction because many of you may have had this same thought.

    In regression, you might have y = 2x

    The interpretation here is "if x increases by 2 units, y increases by 4 units" this is a linear relationship.

    Here you have a nonlinear relationship, log(y) = 2*log(x).

    In fact, this is equivalent to saying that y=x^2 . If you can't see that, google "rules of logs", it's the very first one listed by google. The relationship is not the same as you are used to seeing in regression. The percentage change in x will always depend on the value of x. So if x is 10 and y is 100, then a one unit increase in x (which corresponds to a 10% increase) would yield a 10 unit increase in y (because 10 is 10% of 100).

    The correct answer to this question should be: for changes <20%, the interpretation is as follows: a percentage change in x will effect the same percentage change in y, multiplied by any coefficient on the log(x).

11. Possible loss of interpretation, change of scale (eg: when you invert values, you'll inflate values between 0 and 1), might not apply to all values (eg: cannot invert 0), what happens if you have negative values (eg. cannot take logs of negative values).