

Name: _____

This quiz totals 20 points. The quiz is open-note, but not open-neighbor. Relax and try to do the best you can.

1. (5 points) Explain the difference between a continuous and discrete variable. From a modeling standpoint, why do we care about this difference?

Solution: A continuous variable is one which can take on any value over a continuous range—typically \mathbb{R} (the real number line) or \mathbb{R}^+ (the positive real numbers). A discrete variable is one which can only take on a (typically small) finite set of values.

Within discrete variables, we have three groups:

1. binary (0/1)
2. categorical (0,1,...,K)
3. count (non-negative integers)

2. (5 points) Explain the difference between Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR). Which case will result in imputed values that are most plausible?

Solution: We have:

- **Missing Completely At Random:** missing values randomly happen to be missing
- **Missing At Random:** missing values are not MCAR, but are missing in such a way that we could reconstruct them using other observable information in our data set
- **Missing Not At Random:** missing values are missing in part because of some information that is not contained in the data set

MCAR is easiest to resolve, followed by MAR. *MNAR is essentially impossible to resolve unless you happen to know the process governing the missingness.*

3. (5 points) What is the most common use of the word “optimization” in Data Science? Why is optimization important?

Solution: “Optimization” typically means “finding the minimum of an objective function.” It’s important because it is required to estimate any kind of statistical model!

4. (5 points) Explain what a closed-form solution is. What fraction of objective functions in data science have a closed-form solution?

Solution: A closed-form solution is a solution to a problem that can be expressed analytically (i.e. with a written mathematical formula). Its opposite is a **numerical solution**, which requires a computer to iteratively solve.

As luck would (not) have it, only a very small number of objective functions have a closed-form solution. (Basically, just OLS and a handful of others.)