MA117 - Worksheet 1
Data
March 2, 2021 - Week 1, Tuesday

Problem 1. Here is a data matrix displaying some of the data collected in a survey designed to study smoking habits of Colorado residents.

	sex	age	weight	monthlyIncome	smokeAmount
1	Female	42	120	under $2000	12 cig/day
2	Male	33	150	$2000 to $3000	0 cig/day
3	Female	27	110	$5000 to $6000	0 cig/day
⋮	⋮	⋮	⋮	⋮	⋮

How many variables are in this data matrix? Classify each of the variables as numerical (continuous or discrete?) or categorical (ordinal or nominal?).

Problem 2. A political scientist is interested in the effect of government type on economic development. She wants to use a sample of 30 countries evenly represented among the Americas, Europe, Asia, and Africa to conduct her analysis. What type of sampling strategy should she use to ensure that countries are selected from each region of the world — simple random sampling, cluster sampling, stratified sampling, or multistage sampling? Explain your answer.

Problem 3. Participants in a study report whether or not they used a smart phone before bed and then rate their quality of sleep as either "very poor," "poor," "average," "good," or "very good." A statistician analyzes the data and determines that the two variables are associated.

(a) Determine whether the two variables are ordinal or nominal.
(b) Is it reasonable to conclude from this that smart phone usage before bed causes a change in quality of sleep?
(c) How might you design a study that can be used as evidence for a causal relationship?

Problem 4. Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, "Do you find that your work schedule makes it difficult for you to spend time with your kids after school?" Of the parents who replied, 85% said "no." Is it reasonable to conclude that a majority of parents have no difficulty spending time with their kids after school? Explain your answer.

Problem 5. A study compared percent of internet users and life expectancy at birth in 208 countries and found a positive association. What is a possible confounding variable that might explain this relationship?

Problem 6. Here is some summary data about the success/failure of space launches by different types of launching agencies.

	1957–1999		2000–2018	
	Failure	Success	Failure	Success
Private	13	295	10	562
State	281	3751	33	711
Startup	-	-	5	65

Describe a data matrix that one might use to create this summary table. How many variables are involved? What type is each variable? What is the total number of observations in the data matrix?

Problem 7. You would like to determine if your classmates prefer the taste of regular Coke or Diet Coke. Outline a design for this study.

Problem 8. In 1936, Landon sought the Republican presidential nomination opposing the re-election of FDR. The Literary Digest polled about 10 million Americans and got responses from about 2.4 million. The magazine had surveyed its own readers, registered automobile owners, and registered telephone users. The poll showed that Landon would likely be the overwhelming winner and FDR would get only 43% of the votes.

Election result: FDR won, with 62% of the votes.

The magazine was completely discredited because of the poll and was soon discontinued. What might have gone wrong?