# STAT22000 Winter/Summer 2020 Homework 1

All page, section, and exercise numbers below refer to the course text (*OpenIntro Statistics*, 3rd edition, by Diez, Barr, and Cetinkaya-Rundel.).

**Reading**: Section 1.2, 1.6 (Skip Section 1.6.8)
**Problems for Self-Study** : (Do Not Turn In)

- Exercise 1.5, 1.39, 1.45, 1.47, 1.49, 1.51, 1.53, 1.57 on p.56-69
- Answers can be found at the end of the book (p.405-407).
- Self-study problems are as important as Turn-In problems. We don't require submission because we think you can learn from those problems by doing them yourself and checking the answers, without grading feedbacks. If having questions about those problems, you are welcome to ask the instructor or TAs.

**Problems to Turn In**: due **midnight on Friday, June 26, on Gradescope**.

1. Refer to the description of the Aircraft-Wildlife Collisions data at

   `https://www.openintro.org/data/index.php?data=birds`

   (a) What is a case in this data set?
   (b) Determine whether each of the five variables: `ac_mass`, `effect`, `num_engs`, `height`, and `bird_struck`, is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is nominal or ordinal.

2. Suppose that the cases in a study are the purchases that Yibi made on amazon.com in the past 12 months. Identify each of the following is a variable or not a variable. If it is a variable, determine whether it is numerical or categorical.

   (a) How much did Yibi spend on those purchases in total?
   (b) Was the purchase shipped to Yibi or to someone else?
   (c) Did Yibi spend more on purchases sent to others than on purchases sent to herself?

3. An investigator has a data file showing family incomes for 1,000 subjects in a certain study. The minimum \$5,800 a year to \$98,600 a year. By accident, the highest income in the file gets changed to \$986,000.

   (a) Does this affect the mean? If so, by how much?
   (b) Does this affect the median? If so, by how much?

4. Below are the final exam scores of 25 introductory statistics students.

   42, 53, 63, 76, 76, 78, 80, 85, 86, 86, 87, 87, 88, 88, 89, 89, 90, 91, 92, 94, 95, 95, 96, 96, 97

   Below are the five number summary.

   | Min | Q1 | Median | Q3 | Max |
   |---|---|---|---|---|
   | 42 | 80 | 88 | 92 | 97 |

   (a) Use the 1.5 IQR rule to identify outlier(s) if any.
   (b) Please make a boxplot for the data manually (not by R) and indicate on the plot the values the boundaries and the middle line of the box represent respectively, and show how you determine the values the two whiskers extend to.

5. For each of the following, state whether you expect the distribution to be symmetric, right skewed, or left skewed. Explain your reasoning.

   (a) Housing prices in a country where 25% of the houses cost below $350,000, 50% of the houses cost below $450,000, 75% of the houses cost below $1,000,000 and there are a meaningful number of houses that cost more than $6,000,000.

   (b) Number of alcoholic drinks consumed by college students in a given week. Assume that more than half of these students don't drink since they are under 21 years old, and only a few drink excessively.

   (c) The average on a history exam (scored out of 100 points) was 85, with a standard deviation of 15.