Assignment 1

STA4321 - Intro. Mathematical Statistics I

Dr. A Cohen

Summer 2019

Due by Sunday June 23rd at 11:59PM CT. Submission should only be online via Canvas. Solve all problems and submit via Canvas in a single PDF. A typed submission using LaTeX is preferred and has a bonus of 10%**.**

1. Given in Figure 1 is the relative frequency histogram associated with grade point averages (GPAs) of a sample of students:

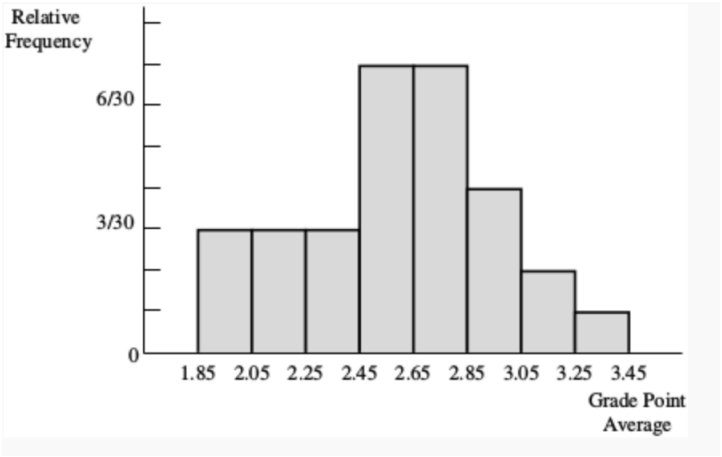

Figure 1: GPA histogram

- Which of the GPA categories identified on the horizontal axis are associated with the largest proportion of students?

- What proportion of students had GPAs in each of the categories that you identified?

- What proportion of the students had GPAs less than 2.65?

2. Write down the formulas of the sample mean, variance, and standard deviation.

   (a) $\overline{y} =$

   (b) $s^2 =$

   (c) $s =$

3. Consider the following sample of 44 observations: 8.9; 12.4; 8.6; 11.3; 9.2; 8.8; 8.8; 6.2; .07; 7.1; 11.8; 10.7; 7.6; 9.1; 9.2; 8.2; 9.0; 8.7; 9.1; 10.9; 10.3; 9.6; 7.8; 11.5; 9.3; 7.9; 8.8; 12.7; 8.4; 7.8; 5.7; 10.5; 10.5; 9.6; 8.9; 10.2; 10.3; 7.7; 10.6; 8.3; 8.8; 9.5; 8.8; 9.4.

   (a) Find the mean and the standard deviation for the data given.

   (b) Calculate the interval for $\overline{y} \pm ks$ for $k \in \{1, 2, 3\}$. Count the number of measurements that fall within each interval and compare this result with the number that you would expect according to the empirical rule for normal approximated data.

4. Resting breathing rates for college-age students are approximately normally distributed with mean 12 and standard deviation 2.3 breaths per minute. What fraction of all college-age students have breathing rates in the following intervals? Use empirical rules.

   (a) 9.7 to 14.3 breaths per minute.

   (b) 7.4 to 16.6 breaths per minute.

   (c) 9.7 to 16.6 breaths per minute.

   (d) Less than 5.1 or more than 18.9 breaths per minute.

5. It has been projected that the average and standard deviation of the amount of time spent

online using the Internet are, respectively, 14 and 17 hours per person per year (many do not use the Internet at all!).

(a) What value is exactly 1 standard deviation below the mean?

(b) If the amount of time spent online using the Internet is approximately normally distributed, what proportion of the users spend an amount of time online that is less than the value you found in question 1?

(c) Is the amount of time spent online using the Internet approximately normally distributed? Why?

6. Weekly maintenance costs for a factory, recorded over a long period of time and adjusted for inflation, tend to have an approximately normal distribution with an average of \$420 and a standard deviation of \$30. If \$450 is budgeted for next week, what is an approximate probability that this budgeted figure will be exceeded?

7. The following results on summations will help us in calculating the variance s^2. For any constant c,

   - $\sum_{i=1}^{n} c = nc$
   - $\sum_{i=1}^{n} cy_i = c\sum_{i=1}^{n} y_i$
   - $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

   (a) Use the above results and show that:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} y_i\Big)^2\right]$$

   (b) Use the the result (a) to find the standard deviation of the following 6 sample measurements: 1, 4, 2, 1, 3, and 3.

8. Suppose a family contains two children of different ages, and we are interested in the gender of these children. Let F denote that a child is female and M that the child is male and let a pair such as FM denote that the older child is female and the younger is male. There are four points in the set of possible observations:

$$S = \{FF, MM, FM, MF\}$$

Let A denote the subset of possibilities containing no males; B, the subset containing two males; and C, the subset containing at least one male. List the elements of A, B, C, $A \cap B$, $A \cup B$, $A \cap C$, $A \cup C$, $B \cap C$, $B \cup C$, and $C \cap \overline{B}$.