ENR211 STATISTICS FOR ENGINEERS

Problem Set 1

---

1 Data Variation

1. (a) An experiment is performed 2-3 times (by you or by other operators at different times). Each gives different values, which value should be reported? And why?
   (b) What is the typical fuel economy $\text{km} \cdot \text{l}^{-1}$ of a car?
   (c) What is the nominal $\text{km} \cdot \text{l}^{-1}$ of your vehicle?
   (d) What kind of data is used to measure electricity units consumed in a month and why?

2. 1 kg standard weight is measured 100 times, what is expected as the measured values. Does one get exactly 100 g each time or something different? Explain

3. Classify each of the following variables as qualitative or quantitative; if quantitative, as discrete or continuous.
   (a) Occupation
   (b) Region of Residence
   (c) Weight
   (d) Height
   (e) Number of automobiles owned

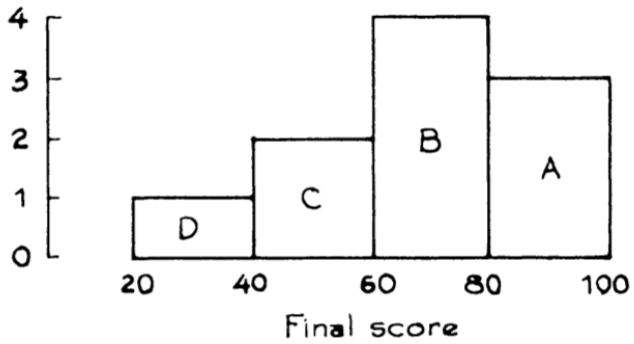
2 Histogram and Cumulative Distribution

1. The histogram below shows the distribution of final scores in a certain class.

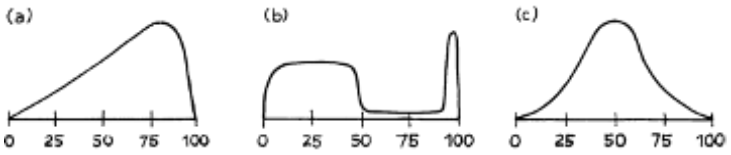
   (a) Which block represents the people who scored between 60 and 80?

   (b) Ten percent scored between 20 and 40. About what percentage scored between 40 and 60?

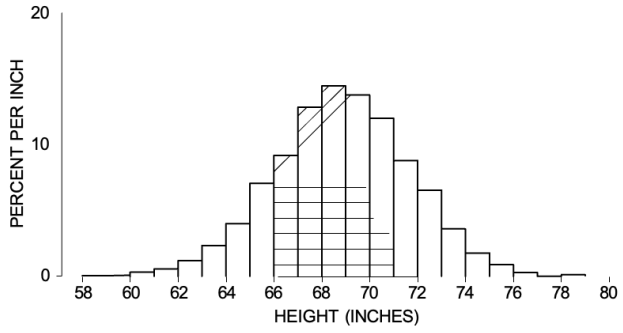
   (c) About what percentage scored over 60?



2. Below are sketches of histograms for test scores in three different classes. The scores range from 0 to 100; a passing score was 50. For each class, was the percent who passed about 50%, well over 50%, or well under 50%?

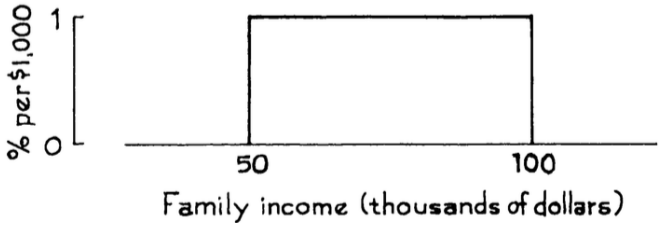


3. The figure below shows a histogram for the heights of a representative sample of men. The shaded area represents the percentage of men whose heights were between ____ and ____. Fill in the blanks.



Source: Data tape supplied by the Inter-University Consortium for Political and Social Research.

4. Someone has sketched one block of a family-income histogram for a wealthy suburb. About what percentage of the families in this suburb had incomes between $90,000 and $100,000 a year?



5. According to the data given below, plot the histogram, calculate the cumulative frequency, relative frequency, plot the Ogive curve, and find its median.

Marks Range	**Frequency**
0–10	4
10–20	6
20–30	10
30–40	15
40–50	15

3 Measures of Central Tendency and Dispersion

1. According to the data given below; find it's Mean, Median, Mode, Range, Standard Deviation and Variance.

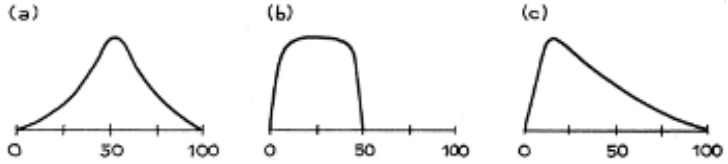
   22, 15, 5, 18, 22, 12, 25, 8, 22, 15, 20

2. Which of the following two lists has a bigger average? Or are they the same? Try to answer without doing any arithmetic.

   (i) 10, 7, 8, 3, 5, 9

   (ii) 10, 7, 8, 3, 5, 9, 11

3. (a) Below are sketches of histograms for three lists. Fill in the blank for each list: the average is around _____. Options: 25, 40, 50, 60, 75.



   (b) For each histogram in exercise 1, is the median equal to the average? or is it to the left? to the right?

4. Here are the heights of the four boys: 150 cm, 130 cm, 165 cm, and 140 cm. Match the heights with the descriptions. A description may be used twice.

   i. Unusually short, **ii.** About average, **iii.** Unusually tall

5. Each of the following lists has an average of 50. For which one is the spread of the numbers around the average biggest? smallest?

   (i) 0, 20, 40, 50, 60, 80, 100

   (ii) 0, 48, 49, 50, 51, 52, 100

   (iii) 0, 1, 2, 50, 98, 99, 100

6. Below are sketches of histograms for three lists. Match the sketch with the description. Some descriptions will be left over. Give your reasoning in each case.

   (i) ave $\approx$ 3.5, SD $\approx$ 1

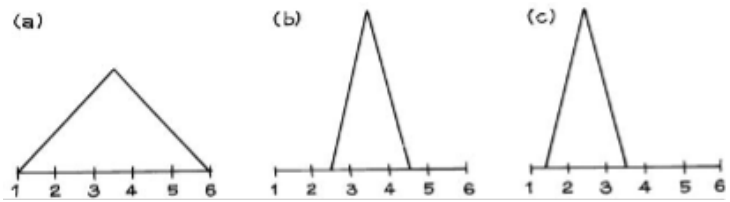
   (ii) ave $\approx$ 3.5, SD $\approx$ 0.5

   (iii) ave $\approx$ 3.5, SD $\approx$ 2

   (iv) ave $\approx$ 2.5, SD $\approx$ 1

   (v) ave $\approx$ 2.5, SD $\approx$ 0.5

   (vi) ave $\approx$ 4.5, SD $\approx$ 0.5



7. Guess which of the following two lists has the larger Standard Deviation. Check your guess by computing the Standard Deviation for both lists.

   (i) 9, 9, 10, 10, 10, 12

   (ii) 7, 8, 10, 11, 11, 13

8. Can the Standard Deviation ever be negative?

9. For a list of positive numbers, can the Standard Deviation ever be larger than the average?

10. Determine whether the following statements are **True** or **False**, and provide a brief explanation for each:

    (a) If you add 7 to each entry on a list, that adds 7 to the average.

    (b) If you add 7 to each entry on a list, that adds 7 to the standard deviation (SD).

    (c) If you double each entry on a list, that doubles the average.

    (d) If you double each entry on a list, that doubles the standard deviation (SD).

    (e) If you change the sign of each entry on a list, that changes the sign of the average.

    (f) If you change the sign of each entry on a list, that changes the sign of the standard deviation(SD).

11. Determine whether the following statements are **True** or **False**. Provide explanations or examples for each:

    (a) The median and the average of any list are always close together.

    (b) Half of a list is always below average.

    (c) With a large, representative sample, the histogram is bound to follow the normal curve quite closely.

    (d) If two lists of numbers have exactly the same average of 50 and the same standard deviation (SD) of 10, then the percentage of entries between 40 and 60 must be exactly the same for both lists.

4 Population vs Sample

1. A survey is carried out at a university to estimate the percentage of undergraduates living at home during the current term. What is the population? the parameter?

2. In the Netherlands, all men take a military pre-induction exam at age 18. The exam includes an intelligence test known as "Raven's progressive matrices," and includes questions about demographic variables like family size. A study was done in 1968 relating the test scores of 18-year-old men to the number of their brothers and sisters. The records of all the exams taken in 1968 were used.

   (a) What is the population? the sample?

   (b) Is there any sampling error? Explain briefly.

3. Polls often conduct pre-election surveys by telephone. Could this bias the results? How? What if the sample is drawn from the telephone book?

4. A survey organization is planning to do an opinion survey of 2,500 people of voting age in the U.S. True or false and explain: the organization will choose people to interview by taking a simple random sample.

5. Two surveys are conducted to measure an advertising campaign's effect for a certain detergent brand. In the first survey, interviewers ask housewives whether they use that brand of detergent. In the second, the interviewers ask to see what detergent is being used. Would you expect the two surveys to reach similar conclusions? Give your reasons.

6. How does the percentage composition of a sample differ from that of the whole population, given that the sample is only part of the population?

5 Random Variable

1. A fair coin is tossed three times. Let X be the random variable representing the number of heads obtained in these three tosses.

   (a) What is the sample space of the experiment?

   (b) What is the probability of each outcome in the sample space?

   (c) Define the random variable X as a function of the outcomes in the sample space.

   (d) Draw the probability mass function (PMF) of X.

2. You roll two dice and observe the sum Z of the two numbers rolled. Let $Z = X + Y$, where X and Y are the numbers rolled on the first and second die, respectively.

   (a) Find the Sample Space and the probability mass function (PMF) of Z.

   (b) Find $P(Z = 8)$.

   (c) Find $P(Z > 5)$.

3. The probability mass function of a random variable X is given by:

$$p(i) = \frac{c\lambda^i}{i!}, \quad i = 0, 1, 2, \ldots$$

   where λ is a positive constant.

   Given this information, find:

   (a) $P(X = 0)$

   (b) $P(X > 2)$

   Hint: The total probability sums up to 1, i.e., $\sum_{i=0}^{\infty} p(i) = 1$.

4. The random variable X is given by:

$$P(X = i) = \log_{10}\left(\frac{i+1}{i}\right) \quad i = 1, 2, 3, \ldots, 9.$$

   (a) Verify that the above is a valid probability mass function by showing that the sum of probabilities equals 1, i.e.,

$$\sum_{i=1}^{9} P(X = i) = 1.$$

   (b) Find $P(X \leq k)$

5. Assume there is a fair coin.

   (a) Let X represent the difference between the number of heads and the number of tails obtained when the coin is tossed n times. What are the possible values of X?

   (b) For $n = 3$, what are the probabilities associated with the values that X can take on?

   (c) Draw the probability mass function (PMF) of X.

6 Probability Distributions

1. Let X be a random variable that follows a binomial distribution with parameters $n = 4$ and $p = \frac{1}{2}$. Draw the probability mass function (PMF) of the random variable X.

2. A fair coin is tossed 5 times. Let X be the random variable representing the number of heads obtained in these 5 tosses.

   (a) Find the probability of obtaining exactly 2 heads.

   (b) Find the probability of obtaining at least 4 heads.

   (c) Find the probability of obtaining at least 1 head but no more than 3 heads.

3. Let Y be a random variable that follows a uniform distribution between the interval [1, 6]. Calculate the probability that $2 \leq Y \leq 4$.

4. Let X be a random variable that follows a binomial distribution with parameters n and p, i.e., $X \sim \text{Binomial}(n, p)$. The probability mass function (PMF) of X is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

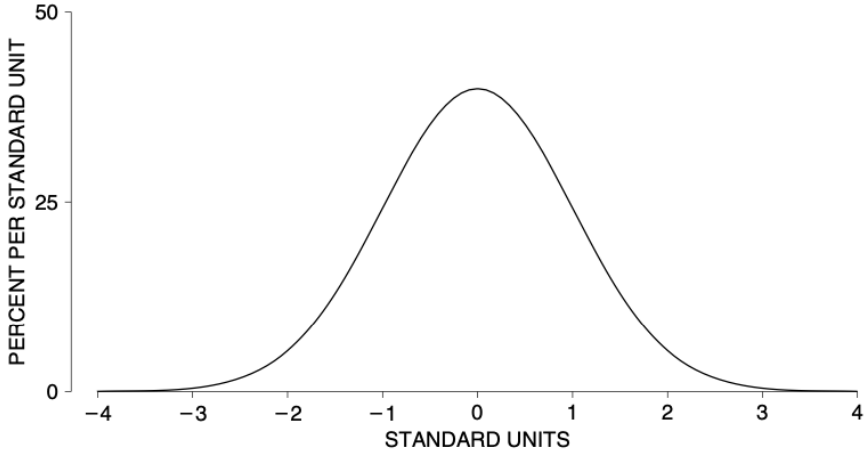
   (a) Prove that the sum of all binomial probabilities is 1, i.e., show that:

$$\sum_{k=0}^{n} P(X = k) = 1.$$

7 Normal Distribution

Features of the Normal Distribution

Figure 1. The normal curve.



Several features of this graph will be important. First, the graph is symmetric about 0: the part of the curve to the right of 0 is a mirror image of the part to the left. Next, the total area under the curve equals 100%.

It will be helpful to find areas under the normal curve between specified values. For instance:

- The area under the normal curve between -1 and $+1$ is about 68%;
- The area under the normal curve between -2 and $+2$ is about 95%;
- The area under the normal curve between -3 and $+3$ is about 99.7%.

---

1. On a certain exam, the average of the scores was 50 and the SD was 10.
   (a) Convert each of the following scores to standard units: 60, 45, 75.
   (b) Find the scores which in standard units are: 0, +1.5, -2.8.

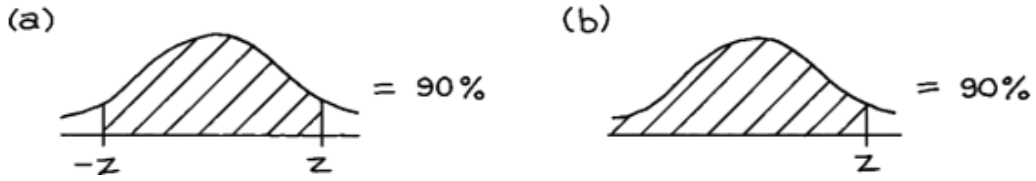
2. (a) Convert each entry on the following list to standard units (that is, using the average and SD of the list): 13, 9, 11, 7, 10.
   (b) Find the average and SD of the converted list.

3. Find the area between 0 and 2 under the normal curve

4. Find the area between -2 and 1 under the normal curve

5. Find the area between -2 and 1 under the normal curve

6. The normal curve is sketched below; solve for z.



7. You are looking at a computer printout of 100 test scores, which have been converted to standard units. The first 10 entries are

   -6.2 3.5 1.2 -0.13 4.3 -5.1 -7.2 -11.3 1.8 6.3

   Does the printout look reasonable, or is something wrong with the computer?