Worksheets

Stephanie Ranft S2459825

October 2, 2019

Statistics 1A
PSBE1-08

1 Week 1

1. Explain what is meant by the term population.
2. Explain what is meant by the term sample.
3. Explain how a sample differs from a population.
4. Explain what is meant by the term sample data.
5. Explain what a parameter is.
6. Explain what a statistic is.
7. Give an example of a population and two different characteristics that may be of interest.
8. Describe the difference between descriptive statistics and inferential statistics. Illustrate with an example.
9. Identify each of the following data sets as either a population or a sample:
   (a) The grade point averages (GPAs) of all students at a college.
   (b) The GPAs of a randomly selected group of students on a college campus.
   (c) The ages of the nine Supreme Court Justices of the United States on January 1,1842 .
   (d) The gender of every second customer who enters a movie theater.
   (e) The lengths of Atlantic croakers caught on a fishing trip to the beach.
10. Identify the following measures as either quantitative or qualitative:
   (a) The 30 high-temperature readings of the last 30 days.
   (b) The scores of 40 students on an English test.
   (c) The blood types of 120 teachers in a middle school.
   (d) The last four digits of social security numbers of all students in a class.
   (e) The numbers on the jerseys of 53 football players on a team.
11. Identify the following measures as either quantitative or qualitative:
   (a) The genders of the first 40 newborns in a hospital one year.
   (b) The natural hair color of 20 randomly selected fashion models.
   (c) The ages of 20 randomly selected fashion models.
   (d) The fuel economy in miles per gallon of 20 new cars purchased last month.
   (e) The political affiliation of 500 randomly selected voters.
12. A researcher wishes to estimate the average amount spent per person by visitors to a theme park. He takes a random sample of forty visitors and obtains an average of $28 per person.
   (a) What is the population of interest?
   (b) What is the parameter of interest?

(c) Based on this sample, do we know the average amount spent per person by visitors to the park? Explain fully.

13. A researcher wishes to estimate the average weight of newborns in South America in the last five years. He takes a random sample of 235 newborns and obtains an average of 3.27 kilograms.

    (a) What is the population of interest?

    (b) What is the parameter of interest?

    (c) Based on this sample, do we know the average weight of newborns in South America? Explain fully.

14. A researcher wishes to estimate the proportion of all adults who own a cell phone. He takes a random sample of 1,572 adults; 1,298 of them own a cell phone, hence 1298/1572≈0.83 or about 83% own a cell phone.

    (a) What is the population of interest?

    (b) What is the parameter of interest?

    (c) What is the statistic involved?

    (d) Based on this sample, do we know the proportion of all adults who own a cell phone? Explain fully.

15. A sociologist wishes to estimate the proportion of all adults in a certain region who have never married. In a random sample of 1,320 adults, 145 have never married, hence 145/1320≈0.11 or about 11% have never married.

    (a) What is the population of interest?

    (b) What is the parameter of interest?

    (c) What is the statistic involved?

    (d) Based on this sample, do we know the proportion of all adults who have never married? Explain fully.

16. What must be true of a sample if it is to give a reliable estimate of the value of a particular population parameter?

2 Week 2

1. The twins Caroline and James have created a table of their school marks, which they got throughout the whole semester in certain subjects.

	Mathematics	Physics	Chemistry	Geography
Caroline	1,2,3,1,1,5,2	3,3,1,1,1,2	1,1,1,3,4,1	2,2,3,1,5,4
James	4,4,1,2,2	2,2,2,2	5,5,4,4,3,4,3,3	1,1,2,1,1,1,2,1

   Calculate the final school mark (as a percentage) of the twins in all subjects, if the range of the school marks is from 1 to 5.

2. The following table contains measured heights of 63 students with the corresponding frequencies:

Height	Frequency	Height	Frequency	Height	Frequency	Height	Frequency
159	1	165	2	170	5	175	2
161	1	166	3	171	6	177	1
162	2	167	2	172	7	178	4
163	1	168	4	173	9	179	2
164	2	169	3	174	5	181	1

   Determine the mean, median, mode, variance and a standard deviation of the student's height.

3. While weighing twenty one-kilogram sugar bags we noted the measured values in kg: 1.00, 1.01, 1.05, 0.99, 0.95, 1.00, 0.98, 0.99, 1.04, 1.06, 0.93, 1.00, 1.03, 0.97, 1.00, 0.99, 1.05, 1.01, 0.94, 1.00. Determine the mean and the variance of the measured weight. Draw a boxplot to represent this data and justify the removal of any outliers (you might need to do this more than once).

4. We measured the height x_i and the weight y_i of ten students; the values are shown in the table below. Find the means of the measured heights and of the measured weights, fill in the table and determine the correlation coefficient between the measured height and weight of the students. There are two methods for calculating the correlation coefficient; describe how you calculate this using both methods and the table below. Using the table, determine the standard deviations of both the heights and weights. What can you determine about the population based on the standard deviations and correlation coefficient?

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	z_{x_i}	z_{y_i}	$z_{x_i} \times z_{y_i}$
1	51	43								
2	54	48								
3	65	56								
4	46	42								
5	58	55								
6	72	54								
7	42	44								
8	69	49								
9	76	55								
10	38	38								
Σ										

5. We measured the flat area of 30 apartments and we have measured the following values in metres squared: 82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3, 61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.

   (a) Create a table of grouped frequency distribution for the 9 classes.

   (b) Construct a histogram of relative frequencies of the flat areas.

   (c) From the specified values estimate the sample mean and the variance.

   (d) From the middle values of the intervals and from their frequencies estimate the sample mean and the variance.

6. The electrical wiring requires cables with a high strength. We examined values for two types of cables:

   1st type: 302, 310, 312, 310, 313, 318, 305, 309, 301, 309, 310, 307, 313, 229, 315, 312, 310, 308, 314, 333, 305, 310, 309, 314

   2nd type: 300, 310, 320, 309, 312, 311, 315, 317, 309, 313, 315, 314, 307, 322, 313, 313, 311, 316, 31, 314, 308, 319, 313, 312

   What can you tell me about the mean strengths of each type, given the information you obtain from the mean, median and mode. Hint: draw a graph which displays the location of each of the three statistics and draw inferences based on the shape.

7. Let X be normally distributed with mean 3 and variance 4. What is $\mathbb{P}(X \geq 6)$? What about $\mathbb{P}(X > 6)$; what can you say about $\mathbb{P}(X = 6)$?

8. Let X be normally distributed with mean 5 and $\mathbb{P}(X \leq 0) = 10\%$. What is the standard deviation of X?

9. Let X be normally distributed with $\mathbb{P}(X \leq 0) = 0.40$ and $\mathbb{P}(X \geq 10) = 0.10$. What is the mean and standard deviation deviation of X?

10. Consider two independent normally distributed random variables X and Y with means 1 and 2, and variances 3 and 4, respectively. What is $\mathbb{P}(X + Y \leq 5)$? Draw a sketch of each distribution, and the distribution of the combined variables $X + Y$. What if the variables were not independent?

For each of the following questions, draw a sketch of the distribution and mark off the mean and the values corresponding to the 68-95-99.7 rule, or any other useful information. For example, if $X \sim \mathcal{N}(50, 5^2)$ (mean 50, sd 5) then

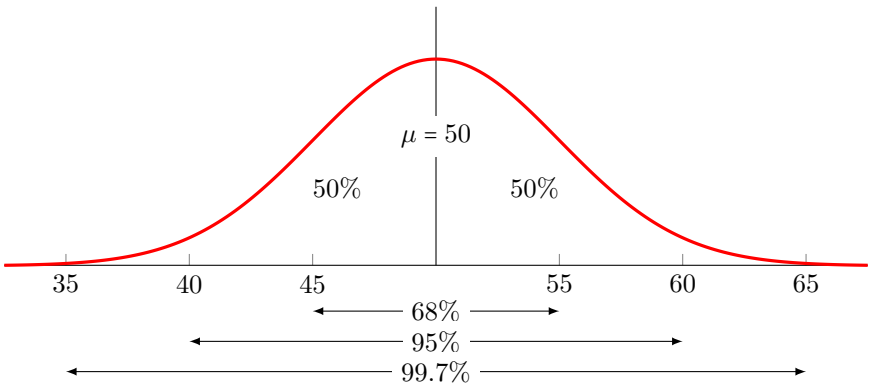

Figure 1: A normally distributed random variable X with mean 50 and variance 25. 50% of the data lies either side of 50, and we can approximate the middle proportion using the 68-95-99.7 rule due to **symmetry about the mean**.

This way, if I am asked what is the probability of having a value greater than 45, I can do some simple addition to find an approximate answer:

$$\mathbb{P}(X > 45) = 50\% + \frac{68\%}{2} = 50\% + 34\% = 84\%.$$

The exact value is 84.1345%, but our estimate is close enough so that we can be sure of our answer when looking at the z-table for the probability.

11. Let X be the IQ of a randomly selected American. Assume $X \sim \mathcal{N}(100, 16^2)$.
    (a) What is the probability that a randomly selected American has an IQ below 90?
    (b) What is the probability that a randomly selected American has an IQ above 140?
    (c) What is the probability that a randomly selected American has an IQ between 92 and 114?

12. Suppose X, the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10.
    (a) The instructor wants to give 15% of the class an A. What cutoff should the instructor use to determine who gets an A?
    (b) The instructor now wants to give 10% of the class an A$^-$. What cutoff should the instructor use to determine who gets an A$^-$?
    (c) The instructor has determined that a grade of 35 or lower is an F. What is the probability of a failing grade?

13. It was found that the mean length of 100 parts produced by a factory was 20.05 mm with a standard deviation of 0.02 mm. Find the probability that a part selected at random would have length
    (a) between 20.03 mm and 20.08 mm;
    (b) between 20.06 mm 20.07 mm;
    (c) less than 20.01 mm;
    (d) greater than 20.09 mm.

14. A company pays its employees an average wage of \$3.25 an hour with a standard deviation of 60 cents. If the wages are approximately normally distributed determine
    (a) the percentage of the workers getting wages between \$2.75 and \$3.69 an hour;
    (b) the minimum wage of the highest 5%.

15. The average life of a certain type of motor is 10 years, with a standard deviation of 2 years. If the manufacturer is willing to replace on 3% of the motors because of failures, how long a guarantee should she offer? Assume that the lives of the motors follow a normal distribution.