Quantitative Methods – I (Statistics)

A. Y. 2022-23

Prof. Lorenzo Cavallo

Exercises

Chapter 1

EXERCISES

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.

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.

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.

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 $1298/1572 \approx .83$ 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 ≈ .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.

Exercise 1.3 Which of the following variables are qualitative, and which are quantitative? Specify which of the quantitative variables are discrete and which are continuous:

Time to travel to work, shoe size, preferred political party, price for a canteen meal, eye colour, gender, wavelength of light, customer satisfaction on a scale from 1 to 10, delivery time for a parcel, blood type, number of goals in a hockey match, height of a child, subject line of an email.

Chapter 2

The data obtained by measuring the age of 21 randomly selected students enrolled in freshman courses at a university could be presented as the data list

18	18	19	19	19	18	22	20	18	18	17
19	18	24	18	20	18	21	20	17	19	

or in set notation as

$$\{18,18,19,19,19,18,22,20,18,18,17,19,18,24,18,20,18,21,20,17,19\}$$

The data obtained by measuring the age of 21 randomly selected students enrolled in freshman courses at a university could be presented as the data list

18	18	19	19	19	18	22	20	18	18	17
19	18	24	18	20	18	21	20	17	19	

or in set notation as

$$\{18,18,19,19,19,18,22,20,18,18,17,19,18,24,18,20,18,21,20,17,19\}$$

The data set of the previous example is represented by the data frequency table

x	17	18	19	20	21	22	24
f	2	8	5	3	1	1	1

1. List all the measurements for the data set represented by the following data frequency table.

x	31	32	33	34	35
f	1	5	6	4	2

1. List all the measurements for the data set represented by the following data frequency table.

x	31	32	33	34	35
f	1	5	6	4	2

$\{31,32,32,32,32,32,33,33,33,33,33,33,34,34,34,34,35,35\}.$

3. Construct the data frequency table for the following data set.

22	25	22	27	24	23
26	24	22	24	26	

3. Construct the data frequency table for the following data set.

22 25 22 27 24 23

26 24 22 24 26

x	22	23	24	25	26	27
f	3	1	3	1	2	1

.

4. Construct the data frequency table for the following data set.

$$\{1,5,2,3,5,1,4,4,4,3,2,5,1,3,2,$$
$$1,1,1,2\}$$

5. A data set contains $n = 10$ observations. The values x and their frequencies f are summarized in the following data frequency table.

x	-1	0	1	2
f	3	4	2	1

a. Construct a frequency bar chart for the data set.
b. Calculate the relative frequencies.

8. A table of some of the relative frequencies computed from a data set is

x	1	2	3	4
f/n	0.3	p	0.2	0.1

The number p is yet to be computed. Finish the table and construct the relative frequency histogram for the data set.

11. During a one-day blood drive 300 people donated blood at a mobile donation center. The blood types of these 300 donors are summarized in the table.

Blood Type	O	A	B	AB
Frequency	136	120	32	12

a. Calculate the relative frequencies.

b. Construct a relative frequency bar chart for the data set.

Exercise 2.2 Consider a variable X describing the time until the first goal was scored in the matches of the 2006 football World Cup competition. Only matches with at least one goal are considered, and goals during the xth minute of extra time are denoted as $90 + x$:

6	24	90+1	8	4	25	3	83	89	34	25	24	18	6
23	10	28	4	63	6	60	5	40	2	22	26	23	26
44	49	34	2	33	9	16	55	23	13	23	4	8	26
70	4	6	60	23	90+5	28	49	6	57	33	56	7	

(a) What is the scale of X?
(b) Write down the frequency table of X based on the following categories: [0, 15), [15, 30), [30, 45), [45, 60), [60, 75), [75, 90), [90, 96).
(c) Draw the histogram for X with intervals relating to the groups from the frequency table.

Exercise 2.4 A university survey was conducted on 500 first-year students to obtain knowledge about the size of their accommodation (in square metres).

j	Size of accommodation (m^2) $e_{j-1} \leq x \leq e_j$	$F(x)$
1	8–14	0.25
2	14–22	0.40
3	22–34	0.75
4	34–50	0.97
5	50–82	1.00

(a) Determine the absolute frequencies for each category.
(b) What proportion of people live in a flat of at least 34 m^2?

2.19 Nixon Corporation manufactures computer monitors. The following data are the numbers of computer monitors produced at the company for a sample of 30 days.

24	32	27	23	33	33	29	25	23	28
21	26	31	22	27	33	27	23	28	29
31	35	34	22	26	28	23	35	31	27

a. Construct a frequency distribution table using the classes 21–23, 24–26, 27–29, 30–32, and 33–35.
b. Calculate the relative frequencies and percentages for all classes.
c. Construct a histogram and a polygon for the percentage distribution.
d. For what percentage of the days is the number of computer monitors produced in the interval 27–29?

2.20 The following data give the numbers of computer keyboards assembled at the Twentieth Century Electronics Company for a sample of 25 days.

45	52	48	41	56	46	44	42	48	53	51	53	51
48	46	43	52	50	54	47	44	47	50	49	52	

a. Make the frequency distribution table for these data.
b. Calculate the relative frequencies for all classes.
c. Construct a histogram for the relative frequency distribution.
d. Construct a polygon for the relative frequency distribution.

Chapter 4

Problem 1

Construct a sample space that describes all three-child families according to the genders of the children with respect to birth order.

Solution:

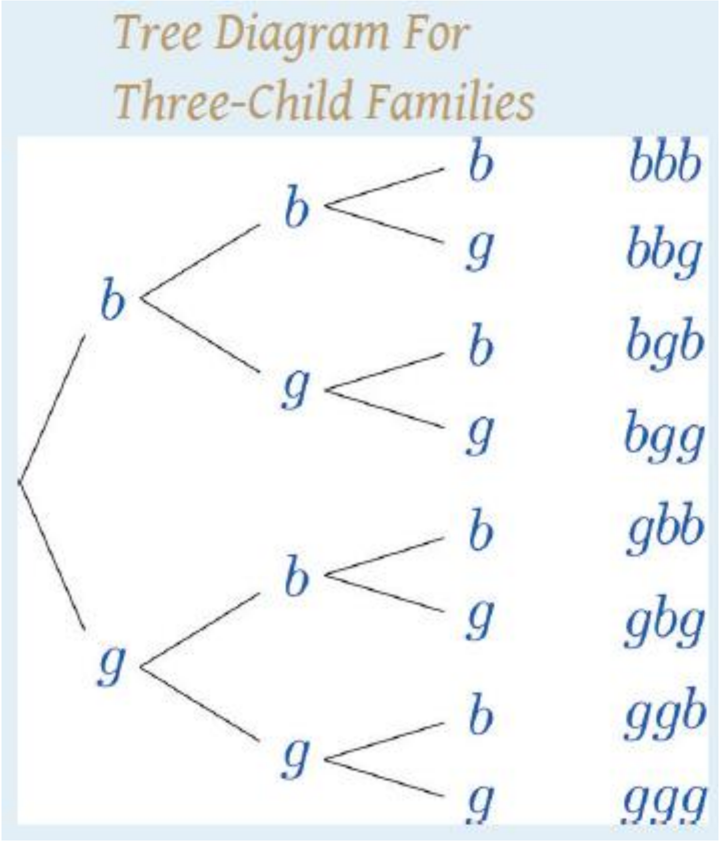
The tree diagram gives a systematic approach.

Hence, the Sample Space is:

$\Omega = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$

Permutation with repetition = $P_{wr\ (n,x)} = P_{wr\ (2,3)} = 2^3$

Nr outcomes for each step = 2 x 2 x 2 = 8



Problem 2

From the sample space that describes all three-child families according to the genders of the children with respect to birth order, identify the outcomes that comprise each of the following events in the experiment of selecting a three-child family at random.

a. At least one child is a girl.
b. At most one child is a girl.
c. All of the children are girls.
d. Exactly two of the children are girls.
e. The first born is a girl.

Solution

$\Omega = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$

a. At least one child is a girl: Event(g ≥ 1): $\{bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$
b. At most one child is a girl:Event(g ≤ 1): $\{bbb, bbg, bgb, gbb\}$
c. All of the children are girls: Event(g = 3): $\{ggg\}$
d. Exactly two of the children are girls: Event(g = 2): $\{bgg, gbg, ggb\}$
e. The first born is a girl: Event(First is a "g"): $\{gbb, gbg, ggb, ggg\}$

a. P(a)=7/8
b. P(b)=4/8=0.5
c. P(c)=1/8
d. P(d)=3/8
e. P(e)=4/8=0.5

Problem 3

You toss a fair coin three times:

a. What is the probability of three heads, HHH?
b. What is the probability that you observe exactly one heads?

Problem 3

You toss a fair coin three times:

a. What is the probability of three heads, HHH?
b. What is the probability that you observe exactly one heads?

We assume that the coin tosses are independent.

a. $P(HHH) = P(H) \cdot P(H) \cdot P(H) = 0.5^3 = \frac{1}{8}$.

b. To find the probability of exactly one heads, we can write

$$\begin{aligned} P(\text{One heads}) &= P(HTT \cup THT \cup TTH) \\ &= P(HTT) + P(THT) + P(TTH) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{3}{8}. \end{aligned}$$

Problem 4

Consider an jar containing three balls, one red, one blue and one green. Two balls are extracted in succession without re-entry.
Determine the probability that, for the extracted pair
1. the first ball drawn is red
2. the second ball extracted is not blue
3. the first ball is red or the second is not blue.

Solution

The number of possible outcomes is 6 from the multiplication rule:

outcomes at the 1st extraction x # outcomes at the 2nd extraction

$$3 \times 2 = 6$$

The Sample Space is $\Omega = \{RB, RG, BR, BG, GR, GB\}$.

Each of these outcomes has probability 1/6

The extractions take place without reintroduction, the probability of extracting a given ball is equal to 1/3 on the first extraction and 1/2 on the second extraction

$$1/3 \times 1/2 = 1/6$$

1. There are two favorable events, {RB, RG}, hence the probability is 2/6.
2. The number of favorable events is 4 {RG, BR, BG, GR}, the probability is 4/6.

Solution

3. We define the events related to the previous points as A = {RB, RG} and B = {RG, BR, BG, GR}.

Hence (A U B) = {RB, RG, BR, BG, GR} has probability 5/6.

By axioms on probability

$$P\,(A\,U\,B)\; =\; P\,(A)\; +\; P\,(B)\; -\; P\,(A\; \cap\; B)$$

with (A ∩ B) = {RB, RG} and P (A ∩ B) = 1/6, we can find the probability also as

$$P(A\,U\,B)\; =\; P\,(A)\; +\; P\,(B)\; -\; P\,(A\; \cap\; B)\; =\; 2/6\; +\; 4/6\; -\; 1/6\; =\; 5/6$$

Problem 5

In a class of 140 students we have 35 students that have failed QM1, 21 that have failed QM2 and 14 that have failed both in QM1 and QM2.

A student is chosen at random.

1. If he failed in QM2, what is the probability that he failed in QM1?
2. If he failed in QM1, what is the probability that he failed in QM2?
3. What is the probability that he failed in QM1 or QM2?

Solution

Let the event A = {the student failed in QM1} and the event B = {the student failed in QM2} , then:

P(A) = 35/140 = 0.25
P(B) = 21/140 = 0.15
P(A ∩ B) = 14/140 = 0.10

1. The probability that a student has failed in QM1, if it is known that failed in QM2, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.10}{0.15} = \frac{2}{3} = 66.67\%$$

Solution

Let the event A = {the student failed in QM1} and the event B = {the student failed in QM2} , then:

P(A) = 35/140 = 0.25
P(B) = 21/140 = 0.15
P(A ∩ B) = 14/140 = 0.10

2. The probability that a student has failed QM2, if it is known that failed in QM1, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.10}{0.25} = \frac{2}{5} = 40\%$$

Solution

Let the event A = {the student failed in QM1} and the event B = {the student failed in QM2} , then:

P(A) = 35/140 = 0.25
P(B) = 21/140 = 0.15
P(A ∩ B) = 14/140 = 0.10

3. The probability that he failed in QM1 or QM2, is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.25 + 0.15 - 0.10 = 0.30 = 30\%$$

Problem 6

In a class we have 10% of students who have had the highest score in statistics.

Of the students who got the highest score in statistics, 40% also had the highest score in math.

If we randomly choose a student, what is the probability that he has passed the highest marks in mathematics and statistics?

Solution

Let the event B = {max score in stats}, then:

P(B) = 0.10

Of the students who got the highest score in statistics, 40% also had the highest score in math. So,

P(A | B) = 0.40

The probability that a student has the max score in math and statistics is

$$P(A \cap B) = P(A|B) \cdot P(B) = 0.40 \cdot 0.10 = 0.04 = 4\%$$

Problem 7

Many diagnostic tests for detecting diseases are not perfectly reliable. The *sensitivity* of a test is the probability that the test will be positive when administered to a person who has the disease. The higher the *sensitivity*, the greater the detection rate and the lower the false negative rate.
Suppose the *sensitivity* of a diagnostic procedure to test whether a person has a particular disease is 92%. A person who actually has the disease is tested for it using this procedure by two independent laboratories.

a. What is the probability that both test results will be positive?
b. What is the probability that at least one of the two test results will be positive?

Solution

a. Let A_1 denote the event "the test by the first laboratory is positive" and let A_2 denote the event "the test by the second laboratory is positive." Since A_1 and A_2 are independent,

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = 0.92 \times 0.92 = 0.8464$$

b. Using the Additive Rule for Probability and the probability just computed,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.92 + 0.92 - 0.8464$$

Problem 8

In the experiment of selecting a three-child family at random, compute each of the following probabilities, assuming all outcomes are equally likely.

a. The probability that the family has at least two boys.
b. The probability that the family has at least two boys, given that not all of the children are girls.
c. The probability that at least one child is a boy.
d. The probability that at least one child is a boy, given that the first born is a girl.

Solution

$\Omega = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$

$\mathrm{X} = \{\textit{the child is a boy}\}$; $\overline{\mathrm{X}} = \{\textit{the child is a girl}\}$

a. The probability that the family has at least two boys:

$$\mathrm{P}(\mathrm{X} \geq 2) = \frac{4}{8} = 0.5$$

b. The probability that the family has at least two boys, given that not all of the children are girls.

$$\mathrm{P}(\mathrm{X} \geq 2 \mid \overline{\mathrm{X}} \neq 3) = \frac{4}{7} = \frac{4/8}{7/8}$$

Solution

$\Omega = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$

$X = \{the\ child\ is\ a\ boy\}$; $\overline{X} = \{the\ child\ is\ a\ girl\}$

c. The probability that at least one child is a boy.

$$P(X \geq 1) = \frac{7}{8} = 0.875$$

d. The probability that at least one child is a boy, given that the first born is a girl.

$$P(X \geq 1 |\ first\ \overline{X}) = \frac{3}{4} = 0.75$$