(1)Please collect a population or a sample data set, and get

(1.1) stem & leaf diagram

(1.2) histograms with 3 different bins (explain how you choose the width of bins)

(1.3) box plot;

Find the mean, mode, median, IQR, (extreme) outliers and standard deviation respectively.

Product *RBrec* is supposed to be a rectangle disk with a red side of 7 cm and a blue side of 5 cm. 50 *RBrecs* are collected and their length of red sides are recorded in cm (account to 2 decimal places) respectively. Let *X* denote the length of red side of *RBrec*.

Since the sample data set is large, Python is used to help with the calculations. Corresponding code is shown here:

import numpy as np

import statistics as stat

N=50 #size of the sample set

x\_mean=np.mean(a) #mean of the sample set

x\_mode=stat.mode(a) #mode of the sample set

x\_per=np.percentile(a, (25, 50, 75), interpolation='midpoint')

IQR=x\_per[2]-x\_per[0] #median,q1,q2 and IQR of the sample set

x\_std=(sum((a-x\_mean)\*\*2)/(N-1))\*\*0.5 # standard deviation of the sample set

print(x\_mean,x\_mode,x\_per[1],x\_std,IQR)

print(a[a>x\_per[2]+1.5\*IQR])

print(a[a>x\_per[2]+3\*IQR])

print(a[a<x\_per[0]-1.5\*IQR])

print(a[a<x\_per[0]-3\*IQR]) #looking for outliers

For this data set, it has:

There are no outliers in this data set.

(1.1)

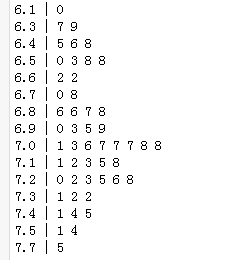


Fig1.1 Stem-leaf diagram of *RBrec*’s length of red side

(1.2)

(1.2.1)

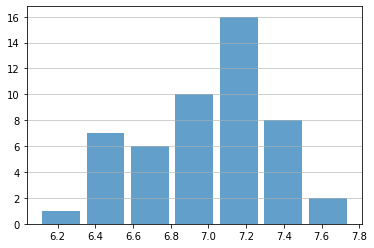


Fig 1.2.1: histogram: equal width with number of bins

(1.2.2)

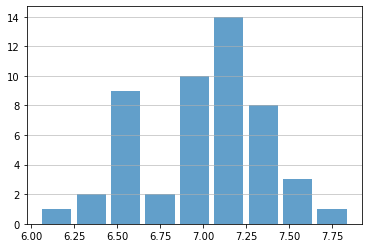


Fig 1.2.2: histogram: equal width but with 9 bins

(1.2.3)

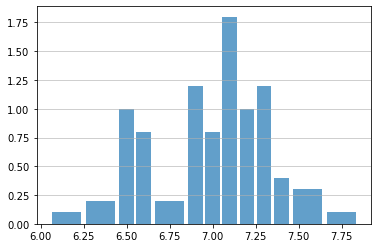


Fig 1.2.3: histogram: unequal width with a shorter bin width around denser area of the data

Fig 1.2.1-1.2.3 show three different histograms of variable *X*.

Fig 1.2.1 has number of bins = square root of the number of observations and give a general bell shape graph of the data set.

Fig 1.2.2 has rational bin boundaries (6.05~6.25,6.25~6.45 etc), which might be helpful if we want to investigate interval probability of the lengths. It also gives us a different image of the data set.

Fig 1.2.3 is based on Fig 1.2.2 but with a shorter width in the region where most data falls. And height of bins in Fig 1.2.3 represents density of data instead of its number.

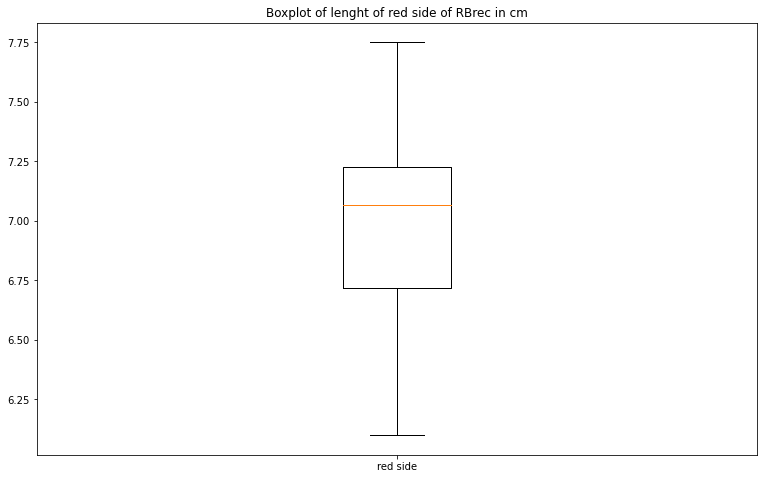
(1.3)

Fig1.3 Boxplot of *RBrec*’s length of red side

(2)Based on the above data set obtained, please try to define two or more random variables and give the joint probability distribution, the marginal distribution(s), and some reasonable conditional probabilities. Please also discuss the relevant covariance and correlation of the variables defined.

Lengths of blue side of these 500 *RBrecs* are recorded together with the lengths of red side. Let *Y* denote the length of blue side of *RBrec*.

7.13, 5.08

Since the sample data set is large, Python is used to help with the calculations. Corresponding code is shown here:

import pandas as pd

A=pd.Series(a)

print(A.value\_counts(normalize=True)) # marginal distribution of variable X

B=pd.Series(b)

print(B.value\_counts(normalize=True)) # marginal distribution of variable Y

for i in range(0,499):

data1[i]=[a[i],b[i]]

Data1=pd.Series(data1)

print(Data1.value\_counts(normalize=True)) # joint probability distribution of variables X and Y

y\_mean=np.mean(b)

y\_std=(sum((a-y\_mean)\*\*2)/(N-1))\*\*0.5

cov=sum((a-x\_mean)\*(b-y\_mean))/(N-1) # covariance between variables X and Y

cor=cov/((x\_std)\*(y\_std)) # correlation between variables X and Y

print(cov,cor)

The joint probability distribution and marginal distributions of variables *X* and *Y* are shown in the table on next page.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X(cm) | Y(cm) | f(X,Y) | X(cm) | Y(cm) | f(X,Y) |
| 6.53 | 5.63 | 0.02 | 7.26 | 5.14 | 0.02 |
| 7.31 | 5.59 | 0.02 | 7.18 | 5.40 | 0.02 |
| 7.22 | 4.76 | 0.02 | 6.62 | 4.99 | 0.02 |
| 7.15 | 4.49 | 0.02 | 7.75 | 4.77 | 0.02 |
| 7.08 | 4.37 | 0.02 | 6.46 | 4.43 | 0.02 |
| 6.93 | 4.71 | 0.02 | 7.07 | 5.16 | 0.02 |
| 7.41 | 5.54 | 0.02 | 6.48 | 5.67 | 0.02 |
| 6.87 | 5.06 | 0.02 | 7.07 | 5.00 | 0.02 |
| 6.95 | 4.49 | 0.02 | 7.54 | 4.45 | 0.02 |
| 7.13 | 5.08 | 0.02 | 7.08 | 5.37 | 0.02 |
| 7.32 | 4.61 | 0.02 | 7.03 | 3.80 | 0.02 |
| 6.58 | 4.55 | 0.02 | 6.37 | 4.88 | 0.02 |
| 7.06 | 4.64 | 0.02 | 6.50 | 5.13 | 0.02 |
| 6.39 | 5.00 | 0.02 | 6.90 | 4.61 | 0.02 |
| 6.58 | 4.94 | 0.02 | 6.10 | 5.64 | 0.02 |
| 7.44 | 4.35 | 0.02 | 7.32 | 4.73 | 0.02 |
| 7.07 | 4.04 | 0.02 | 6.86 | 4.46 | 0.02 |
| 7.51 | 4.93 | 0.02 | 6.88 | 5.73 | 0.02 |
| 7.28 | 4.62 | 0.02 | 7.20 | 5.17 | 0.02 |
| 6.70 | 5.25 | 0.02 | 7.11 | 4.07 | 0.02 |
| 6.99 | 5.17 | 0.02 | 6.86 | 5.22 | 0.02 |
| 7.25 | 4.97 | 0.02 | 6.78 | 4.77 | 0.02 |
| 7.23 | 5.03 | 0.02 | 7.45 | 5.33 | 0.02 |
| 7.01 | 5.28 | 0.02 | 7.12 | 6.34 | 0.02 |
| 6.62 | 4.68 | 0.02 | 6.45 | 4.89 | 0.02 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X(cm) | f(X) | X(cm) | f(X) |  | Y(cm) | f(Y) | Y(cm) | f(Y) |
| 6.10 | 0.02 | 7.07 | 0.06 | 3.80 | 0.02 | 4.99 | 0.02 |
| 6.37 | 0.02 | 7.08 | 0.04 | 4.04 | 0.02 | 5.00 | 0.04 |
| 6.39 | 0.02 | 7.11 | 0.02 | 4.07 | 0.02 | 5.03 | 0.02 |
| 6.45 | 0.02 | 7.12 | 0.02 | 4.34 | 0.02 | 5.05 | 0.02 |
| 6.46 | 0.02 | 7.13 | 0.02 | 4.37 | 0.02 | 5.08 | 0.02 |
| 6.48 | 0.02 | 7.15 | 0.02 | 4.43 | 0.02 | 5.13 | 0.02 |
| 6.50 | 0.02 | 7.18 | 0.02 | 4.45 | 0.02 | 5.14 | 0.02 |
| 6.53 | 0.02 | 7.20 | 0.02 | 4.46 | 0.02 | 5.16 | 0.02 |
| 6.58 | 0.04 | 7.22 | 0.02 | 4.49 | 0.04 | 5.17 | 0.04 |
| 6.62 | 0.04 | 7.23 | 0.02 | 4.55 | 0.02 | 5.22 | 0.02 |
| 6.70 | 0.02 | 7.25 | 0.02 | 4.61 | 0.04 | 5.25 | 0.02 |
| 6.78 | 0.02 | 7.26 | 0.02 | 4.62 | 0.02 | 5.28 | 0.02 |
| 6.86 | 0.04 | 7.28 | 0.02 | 4.63 | 0.02 | 5.33 | 0.02 |
| 6.87 | 0.02 | 7.31 | 0.02 | 4.68 | 0.02 | 5.37 | 0.02 |
| 6.88 | 0.02 | 7.32 | 0.04 | 4.71 | 0.02 | 5.40 | 0.02 |
| 6.90 | 0.02 | 7.41 | 0.02 | 4.73 | 0.02 | 5.54 | 0.02 |
| 6.93 | 0.02 | 7.44 | 0.02 | 4.76 | 0.02 | 5.59 | 0.02 |
| 6.95 | 0.02 | 7.45 | 0.02 | 4.77 | 0.04 | 5.63 | 0.02 |
| 6.99 | 0.02 | 7.51 | 0.02 | 4.88 | 0.04 | 5.64 | 0.02 |
| 7.01 | 0.02 | 7.54 | 0.02 | 4.93 | 0.02 | 5.67 | 0.02 |
| 7.03 | 0.02 | 7.75 | 0.02 | 4.94 | 0.02 | 5.73 | 0.02 |
| 7.06 | 0.02 |  |  | 4.97 | 0.02 | 6.34 | 0.02 |

Some conditional probability:

The Correlation of variable *X* and *Y* is close to 0. If we can assume that variables *X* and *Y* have normal distribution respectively, we can deduce that these two random variables are independent.

(3)

(3.1)Multinomial probability distribution

Suppose a random experiment consists of a series of *n* independent trials and each trial has *k* different possible outcomes corresponding

.

Let random variables the number of trials that result in class 1, class 2,, class *k*, respectively, then they have a **multinomial distribution** and the

joint probability mass function is:

For and .

The marginal probability distribution of is binomial with

Example:

Test results from an electronic circuit board indicate that 50% of board failures are caused by assembly defects, 30% are due to electrical components, and 20% are due to mechanical defects. Suppose that 10 boards fail independently. Let the random variables *X*, *Y*, and *Z* denote the number of assembly, electrical, and mechanical defects among the10 boards.

Random variables *X*, *Y*, and *Z* have a **multinomial distribution** with joint probability mass function:

(a)

(b) is a binomial distribution with Hence

(c) is a binomial distribution with Hence

(3.2)Bivariate Normal distribution

If two random variables *X* and *Y* have a **Bivariate normal distribution**, their joint probability function is:

With parameters ,, and the marginal probability distributionsof *X* and *Y* are normal with means and and standard deviations and , respectively. represents the correlation between *X* and *Y.*

Furthermore the conditional probability distribution of *Y* given is a normal distribution with mean and variance shown below:

And from the parameters of the conditional probability distribution, we can deduce that if , then random variables *X* and *Y* are independent.

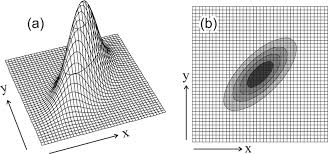


Fig2.1: Illustrations of the correlated bivariate normal distribution

Example:

In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Let *X* and *Y* denote the milliliters of acid and base needed for equivalence, respectively. Assume *X* and *Y* have a bivariate normal distrib-

ution with .

(a)Marginal probability distribution of *X* and *Y*

(b)

(c)Conditional probability distribution of *X* given that

(d)