MONASH University

Information Technology

程序代写代做 CS编程辅导

FIT1006 Business Information Analysis

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

Lecture 3 Graphical Presentation of (Quantitative) Data

Topics covered:

程序代写代做 CS编程辅导

- Types of Data
- Tally & Frequency
- Stem and Leaf Plot
- Histogram
- Visualisation of Data.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

- You can go through pie charts and column graphs in Excel on your own.
- We will cover multivariate data and time series later in the course.

Motivating problem…

程序代写代做 CS编程辅导

- A grocery store war[illegible] analyse the amount spent by their customers. They als[illegible]here might be different types of customers. They ha[illegible] you the sales history of 10 randomly sampled customers.

WeChat: cstutorcs

- Oh, you can't use a calculator or computer…

Assignment Project Exam Help

Email: tutorcs@163.com

- We'll work on this in Lectures 4 & 5 also…

QQ: 749389476

https://tutorcs.com

Introduction

程序代写代做 CS编程辅导

- Different types of data allow us to perform different methods of analysis
- The first step in analysing quantitative data should be to 'see' the data in order to observe the underlying pattern, or distribution, of the data.
- It is the distribution of the data which determines which statistical measures are appropriate.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

程序代写代做 CS编程辅导

https://flux.qa (Feed code: SJ6KGV)

Question 1

I dig up 10 worms in my back garden and measure their length. What type of data do I have?

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

A. Discrete
B. Continuous
C. Qualitative
D. Categorical
E. Something else!

https://flux.qa (Feed code: SJ6KGV)

程序代写代做 CS编程辅导

Question 2

Students complete a survey ranking their satisfaction as high, medium or low. What type of data is that?

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

A. Discrete
B. Continuous
C. Qualitative
D. Categorical
E. Something else!

non-numerical, words

https://flux.qa (Feed code: SJ6KGV)

程序代写代做 CS编程辅导

Question 3

Students complete a survey ranking their satisfaction on a scale 0, 1, 2, …, 10. What type of data is that?

WeChat: cstutorcs

A. Discrete

Assignment Project Exam Help

B. Continuous

Email: tutorcs@163.com

C. Qualitative

QQ: 749389476

D. Categorical

https://tutorcs.com

E. Something else!

Types of Data

程序代写代做 CS编程辅导

- Quantitative
  - Discrete , 2, 3...
  - Continuou , 4.256, 3.999, 11.2, ...
  - *We can calculate summary numerical statistics*
- Ordinal
  - Qualitative eg. agree, neutral, disagree
  - Qualitative eg. Lecturer, Senior Lecturer, Professor
  - *We can calculate statistics based on rank*
- Nominal
  - Categorical eg. red, green, silver, ...
  - *We can calculate statistics based on counts*

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

程序代写代做 CS编程辅导

Motivating Problem

- Data is from the 'Dunnhumby's Shopper Challenge' which recorded the amount spent and date of the transaction at a supermarket in the US over one year.
- See: http://www.kaggle.com/c/dunnhumbychallenge
- I have resampled 'spend' from the original data, using approx 20% of the original observations.
- We will analyse data for 10 groups of shoppers.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Sample Data

程序代写代做 CS编程辅导

- The data below is for Customer #208

53, 16, 66, 10, 77, 25, 17, 44, 37, 25, 24, 62, 3, 50, 16, 18, 5, 31, 29

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

- What can you say about that customer?

Determining class intervals

程序代写代做 CS编程辅导

- Some guidelines:
  - Class intervals lie between possible observations.
  - Express cut-offs at one more decimal point accuracy than data.
  - For n data, $\sqrt{n}$ intervals is usually a good starting point.
  - Sturge's formula: classes = $1 + 3.3\log_{10}n$
  - Choose 'intuitive' intervals: 1, 2, 5, 10, 20, 50, 100 etc.
  - It is more usual to use equal sized intervals.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Sample Data – class intervals

程序代写代做 CS编程辅导

53, 16, 66, 10, 77, 7, 44, 37, 25, 24, 62, 3, 50,
16, 18, 5, 31, 29

19 observations

WeCHat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Smallest = 3

Largest = 77

Intervals ??

Intervals $= \sqrt{19} = 4.4$

Or using Sturge's
$k = 1 + 3.3 \log 19 = 5.2$ classes

$$\text{Intervals} = \frac{\text{Range}}{\text{No. of classes}}$$

$$= \frac{77 - 3}{5.2} = 14.2$$

✓ Intervals (ideally) = 5

程序代写代做 CS编程辅导

Sample Data – Stem and leaf plot

53, 16, 66, 10, 77, 7, 44, 37, 25, 24, 62, 3, 50, 16, 18, 5, 31, 29

stem	leaf
0	3 5
1	0 6 6 7 8
2	4 5 5 9
3	1 7
4	4
5	0 3
6	2 6
7	7

2
7
(4)
8
6
5
3
1

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Sample Data – Stem and leaf plot

程序代写代做 CS编程辅导

53, 16, 66, 10,

17, 44, 37, 25,

3, 50, 16, 18, 5, 31, 29.

```
Stem and Leaf Plot of Variable:
RSPEND, N = 19

Minimum      :  3.000
Lower Hinge  : 16.500
Median       : 25.000
Upper Hinge  : 47.000
Maximum      : 77.000

 0    35
 1 H  06678
 2 M  4559
 3    17
 4 H  4
 5    03
 6    26
 7    7
```

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

- Output from SYSTAT on the RHS
- What can you say about that customer?

Sample Data – Histogram

程序代写代做 CS编程辅导

53, 16, 66, 10,

17, 44, 37, 25,

3, 50, 16, 18, 5, 31, 29.

WeChat: cstutorcs

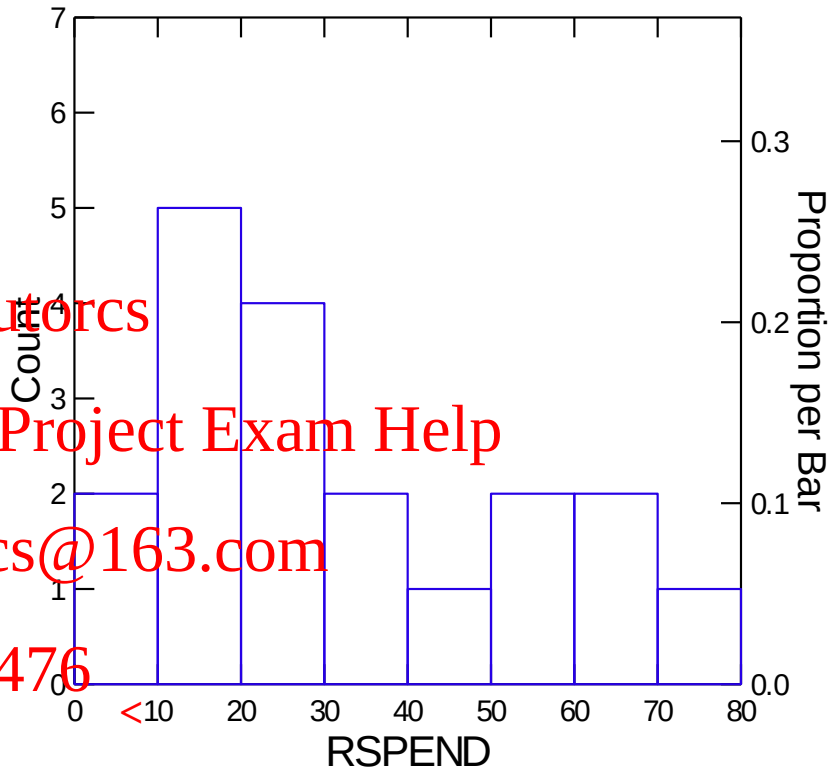
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

- Output from SYSTAT on the RHS




```
 0   35
 1 H 06678
 2 M 4559
 3    17
 4 H 4
 5    03
 6    26
 7
```

Sample Data – Dotplot

程序代写代做 CS编程辅导

53, 16, 66, 10,

17, 44, 37, 25,

3, 50, 16, 18, 5, 31, 29.

```
  0   35
  1 H 06678
  2 M 4559
  3   17
  4 H 4
  5   03
  6   26
  7   7
```

WeChat: cstutorcs

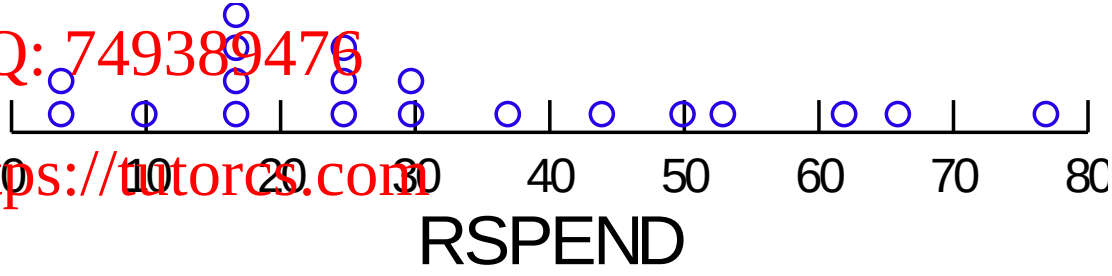
Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

- Output from SYSTAT



Motivating Problem

程序代写代做 CS编程辅导

- Working in grou[illegible], each group will draw a stem and leaf p[illegible]one of the 10 customers. Your customer is based on the first letter of your last name indicated in the worksheet.

WeChat: cstutorcs

Assignment Project Exam Help

- Describe the shape of the distribution of data.

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

Motivating Problem – SYSTAT

程序代写代做 CS编程辅导

```
ID40(0), N = 13

  0   267
  1 M 3447
  2
  3   9
  4 H 05
  5   4
  6   13

ID79(1), N = 10

  0 H 01233
  0 M 7
  1 H 11
  1   5
  2   3

ID119(2), N = 21

  0   22223
  0 H 68
  1
  1   6
  2 M 000012
  2   6
  3 H 002
  3
  4   1
  4   5
  5
  5   5
```

```
DI123(3), N =

  7 H 045
  8
  9   4
 10   1
 11   0

ID134(4), N = 66

  0   022
  0   56677789
  1 H 0012234
  1   5555567788889
  2 M 111223344
  2   556678
  3   22
  3 H 58999
  4   0114
  4   68
  5   0014
  5
  6
  6
  7   2
* * * Outside Values * * *
  9   69
 12   1
```

```
ID140(5), N = 32

  0   134699
  1 H 129
  2   3
  3   4
  4   27
  5 M 224467
  6   235
  7 H 33
  8   26
  9   04
 10   357
 11   4

ID148(6), N = 49

  0   111
  0   22
  0   4455555
  0 H 6666667777
  0 M 8899
  1   00
  1   2233
  1   4
  1   7
  1   8
  2 H 001
  2   223
  2   4
  2   6
  2   899
  3
  3   2
* * * Outside Values * * *
  6   6
  6   6
```

```
ID149(7), N = 11

  0   45
  1   4
  2 H 89
  3 M 6
  4   0
  5 H 44
  6   9
  7   7

ID168(8), N = 29

  0   24
  0   6779
  1 H 444
  1   66688
  2 M 0122
  2   9
  3 H 0004
  3   6
  4   3
  4
  5   44
* * * Outside Values * * *
  6   8
 14   1

ID177(9), N = 10

  4   9
  5
  6 M 1334
  7   3
  8   1
  9 H 6
 10   79
```

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Describe the different types of customers…

Visualising Data

程序代写代做 CS编程辅导

- Why do we want to [illegible] data?
  - Visual inspecti[illegible] fastest way to get an overview of data.
  - Visual inspection enables a description of the distribution of the data to be made.
  - The distribution of data determines which statistics are appropriate.
  - To make comparisons between data from different groups.

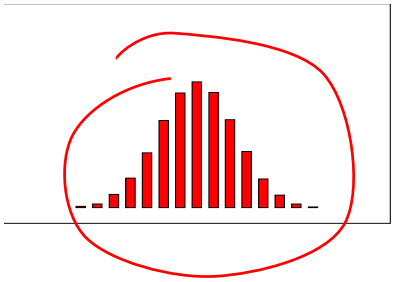
WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Comparing Distributions

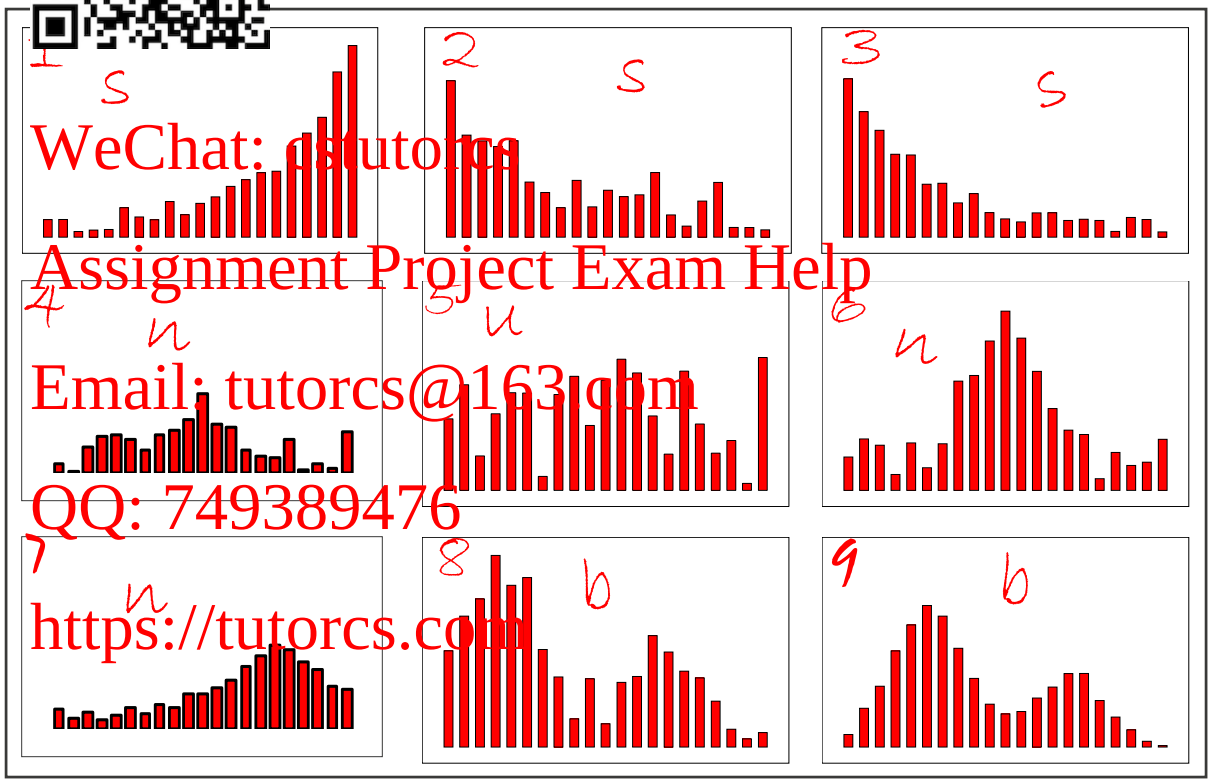
程序代写代做CS编程辅导

Histograms can take other distributions

Normal Dist.



Some ***skewed, bimodal, and uniform*** distributions



WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com

Key Ideas

程序代写代做 CS编程辅导

- You should be able to construct a tally, histogram, cumulative frequency plot and stem-and-leaf plot of a data set.
- You should be able to calculate an appropriate class interval, determine cut-offs and mid-points.
- Get into the habit of always visualising a data set before you undertake any analysis.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

程序代写代做 CS编程辅导

Reading/Questions (Selvanathan)

- Reading: Graphical [illegible]ve Methods
  - 7th Ed. Section [illegible], 4.1
- Questions: Graphical Descriptive Methods
  - 7th Ed. 2.4, 2.5, 2.9, 4.4, 4.5, 4.6, 4.10.
- Tutorial 2 Questions.

WeChat: cstutorcs
Assignment Project Exam Help
Email: tutorcs@163.com
QQ: 749389476
https://tutorcs.com