



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

Lecture 5

QQ: 749389476

Descriptive Statistics – Introduction to
EXCEL and SYSTAT

<https://tutores.com>

Topics covered:

程序代写代做 CS编程辅导

- Calculating descriptive statistics with EXCEL and SYSTAT.
- Comparing groups
- Visualising data
- Using appropriate statistics
- Describing data



WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Learning Objectives

程序代写代做 CS编程辅导



- This lecture is about how to characterise a data set using some summary statistics
- A typical problem that could be answered with the techniques covered today is: describe the differences between the two data sets A and B below?

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

A ← X X XXXX XX XXXXXXXX X XXXX XX X X →

QQ: 749389476

B ← XX X XX X XXXX →

<https://tutorcs.com>

程序代写代做 CS编程辅导

Motivating problem...



- A grocery store wants to analyse the amount spent by their customers. They also think there might be different types of customers. They have given you the sales history of 10 randomly sampled customers.
- Data is from the Kaggle '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 the original data, using approx 20% of the original observations.
- We will use the data for 10 groups of shoppers.

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



- Working in groups using the data for Customer 3 (shown on the right) do the following:

- Draw a stem and leaf plot.
- Calculate the quartiles using the quick method.
- Calculate Q1 using $q = (n+1) \frac{Q}{4}$
- Calculate a 10% trimmed mean.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476


<https://tutorcs.com>

22
18
13
37
14
74
70
62
75
16
21
33
11
101
114
5
94
2
33
10

程序代写代做 CS编程辅导

Sample Data – Stem and leaf plot

22
18
13
37
14
74
70
62
75
16
21
33
11
101
114
5
94
2
33
10

Stem & leaf		
		
25		
013468		
2		12
3		337
4		
5		
6		2
7		045
8		
9		4
10		1
11		4

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

<https://flux.qa> ^{程序代写代做CS编程辅导} (Feed code: SJ6KGV)

Question 1



For Customer 3 the Quick Method, Q1 =

A. 13.00

B. 13.25

C. 13.50

D. 13.75

E. 14.00

A 13.00

(2) 4.35%

B 13.25

Assignment Project Exam Help (18) 39.13%

C 13.50

Email: tutorcs@163.com (20) 43.48% ✓

D 13.75

QQ: 749389476 (4) 8.7%

E 14.00

<https://tutorcs.com> (2) 4.35%

2	5	10	11	13	14	16	18	21	22	33	33	37	62	70	74	75	94	101	114
---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

4 x 5 = 20 --> n = 5

Q1

Q2

Q3

<https://flux.qa>

程序代写 代做 CS 编程辅导
(Feed code: SJ6KGV)

Question 2



■ For Customer 3

$$q = (n + 1) \frac{Q}{4}, Q_1 =$$

2	5	10	11	13	14	16	18	21	22	33	33	37	62	70	74	75	94	101	114
---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

WeChat: cstutorcs

A. 13.00

B. 13.25

C. 13.50

D. 13.75

E. 14.00

Assignment Project Exam Help

Email: tutorcs@163.com

$Q = x_q + r(x_{q+1} - x_q)$ when q is non-integer
QQ: 749389476

<https://tutorcs.com>

$Q_1 \rightarrow 5^{\text{th}} \text{ value} + 0.25 (6^{\text{th}} \text{ value} - 5^{\text{th}} \text{ value})$
 $Q_1 \rightarrow 13 + 0.25 (14 - 13) = 13.25$

<https://flux.qa> (程序代写代做CS编程辅导)
(Feed code: SJ6KGV)

Question 3



For Customer 3, the mean, using the 10% trimmed mean:

A. 30.15

B. 37.69

C. 39.39

D. 41.25

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

2	5	10	11	13	14	16	18	21	22	33	33	37	62	70	74	75	94	101	114
--------------	--------------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----------------	----------------

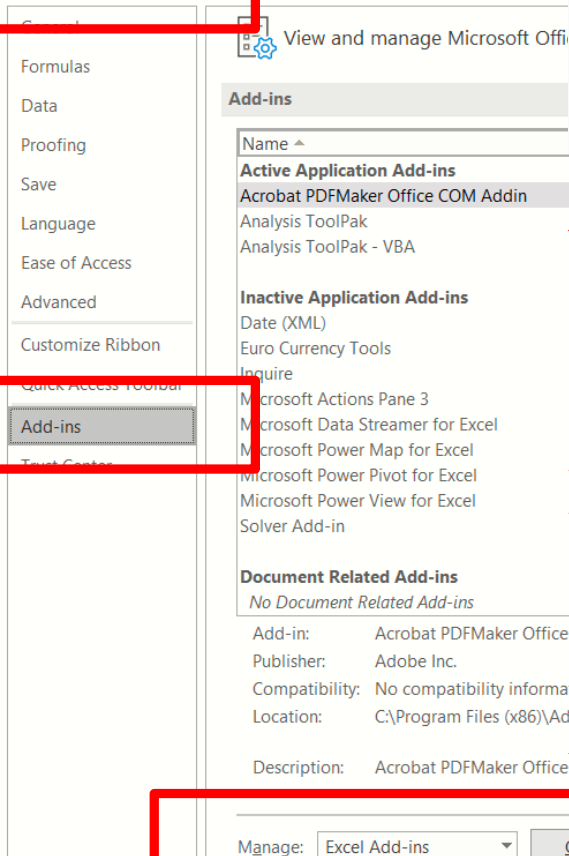
10% trimmed mean : $(10 + 11 + 13 + 14 + 16 + 18 + 21 + 22 + 33 + 33 + 37 + 62 + 70 + 74 + 75 + 94) / 16 = 37.69$

Analysis Tools – Excel Add-ins

程序代写代做 CS编程辅导



Excel Options



Add-ins

Add-ins available:

- ☒ Analysis ToolPak
- ☒ Analysis ToolPak - VBA
- ☐ Euro Currency Tools
- ☐ Solver Add-in

OK

Cancel

Browse...

Automation...

WeChat: estutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Analysis ToolPak

Provides data analysis tools for statistical and engineering analysis

OK

Cancel

Data -> Data Analysis

程序代写代做 CS编程辅导



Excel interface showing the Data Analysis tool selected in the ribbon. The Data Analysis dialog box is open, displaying a list of analysis tools. The 'Data' range is highlighted in the spreadsheet.

Data Analysis Tools:

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

Data Range:

	A	B	C	D
1	Data		Data	
2	22			
3	18	Mean		41.25
4	13	Standard Error		7.866275451
5	37	Median		33
6	14	Mode		33
7	74	Standard Deviation		35.17905328
8	70	Sample Variance		1237.565789
9	62	Kurtosis		-0.71349599
10	75	Skewness		0.801922867
11	16	Range		112
12	21	Minimum		2
13	33	Maximum		114
14	11	Sum		825
15	101	Count		20
16	114			
17	5			
18	94			
19	2			
20	33			
21	10			

WeChat: cstutorcs

Assignment Project Exam Help

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),



7 H 045

8

9 4

10 1

11 4

ID134(4), N = 66

```
0 022
0 56677789
0 112233
1 555666788889
2 M 111223344
2 556678
3 22
3 H 18999
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 449556
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
4 6
9 6
```

*** Outside Values ***

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...

Motivating Problem - SYSTAT

程序代写代做 CS编程辅导



Stem and Leaf Plot of Variable:

Minimum : 2.000
Lower Hinge : 13.000
Median : 17.000
Upper Hinge : 45.000
Maximum : 63.000

Stem and Leaf Plot of Variable: ID79(1), N = 10

Minimum : 5.000
Lower Hinge : 25.000
Median : 57.500
Upper Hinge : 115.000
Maximum : 239.000

Stem and Leaf Plot of Variable: ID119(2), N = 21

Minimum : 2.000
Lower Hinge : 6.000
Median : 20.000
Upper Hinge : 30.000
Maximum : 55.000

Stem and Leaf Plot of Variable: ID113(3), N = 20

Minimum : 2.000
Lower Hinge : 13.500
Median : 27.500
Upper Hinge : 72.000
Maximum : 114.000

Stem and Leaf Plot of Variable: ID134(4), N = 66

Minimum : 0.000
Lower Hinge : 13.000
Median : 21.500
Upper Hinge : 39.000
Maximum : 121.000

Stem and Leaf Plot of Variable: ID140(5), N = 32

Minimum : 1.000
Lower Hinge : 15.500
Median : 54.000
Upper Hinge : 77.500
Maximum : 114.000

Stem and Leaf Plot of Variable: ID148(6), N = 49

Minimum : 1.000
Lower Hinge : 6.000
Median : 9.000
Upper Hinge : 20.000
Maximum : 96.000

Stem and Leaf Plot of Variable: ID149(7), N = 11

Minimum : 1.000
Lower Hinge : 21.000
Median : 36.000
Upper Hinge : 54.000
Maximum : 77.000

Stem and Leaf Plot of Variable: ID168(8), N = 29

Minimum : 2.000
Lower Hinge : 14.000
Median : 20.000
Upper Hinge : 30.000
Maximum : 141.000

Stem and Leaf Plot of Variable: ID177(9), N = 10

Minimum : 49.000
Lower Hinge : 63.000
Median : 68.500
Upper Hinge : 96.000
Maximum : 109.000

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Or use Excel...

程序代写代做 CS编程辅导

- Descriptive Statistics (Meaning up).



	ID40(0)	ID79(1)	ID133(3)	ID134(4)	ID140(5)	ID148(6)	ID149(7)	ID168(8)	ID177(9)	
Mean	28.85	80.90	20.14	41.25	27.38	51.94	14.18	37.27	27.41	76.60
Standard Error	6.14	23.60	3.32	7.87	2.81	6.16	2.21	7.38	5.02	6.60
Median	17.00	57.50	20.00	27.50	21.50	54.00	9.00	36.00	20.00	68.50
Mode	14.00	#N/A	20.00	33.00	15.00	73.00	6.00	54.00	16.00	63.00
Standard Deviation	22.12	74.63	15.19	35.18	22.82	34.85	15.45	24.49	27.01	20.86
Sample Variance	489.47	5569.66	230.83	1237.57	520.76	1214.82	238.74	599.82	729.75	435.16
Kurtosis	-1.55	0.80	-0.33	-0.71	5.23	-1.11	16.00	-0.97	10.93	-1.08
Skewness	0.38	1.09	0.58	0.80	2.03	0.08	3.40	0.16	2.91	0.58
Range	61.00	234.00	53.00	112.00	121.00	118.00	95.00	73.00	139.00	60.00
Minimum	2.00	5.00	2.00	2.00	0.00	1.00	1.00	4.00	2.00	49.00
Maximum	63.00	239.00	55.00	114.00	121.00	114.00	96.00	77.00	141.00	109.00
Sum	375.00	809.00	423.00	825.00	1807.00	1662.00	695.00	410.00	795.00	766.00
Count	13.00	10.00	21.00	20.00	66.00	32.00	49.00	11.00	29.00	10.00

<https://tutorcs.com>

Or SYSTAT...

程序代写代做 CS编程辅导

Summary Statistics



	ID40 (0)	ID79 (1)	ID134 (4)	ID140 (5)	ID148 (6)	ID149 (7)	ID168 (8)	ID177 (9)		
N of Cases	13	10	21	20	66	32	49	11	29	10
Minimum	2.000	5.000	2.000	2.000	0.000	1.000	1.000	4.000	2.000	49.000
Maximum	63.000	239.000	55.000	114.000	121.000	114.000	96.000	77.000	141.000	109.000
Median	17.000	57.500	20.000	27.500	21.500	54.000	9.000	36.000	20.000	68.500
Arithmetic Mean	28.846	80.900	20.143	41.250	27.379	51.938	14.184	37.273	27.414	76.600
Standard Deviation	22.124	74.630	15.193	35.179	22.820	31.847	15.451	24.491	27.014	20.860
Method = CLEVELAND										
1 of 4	11.500	25.000	5.250	13.500	13.000	15.500	5.750	17.500	14.000	63.000
2 of 4	17.000	57.500	20.000	27.500	21.500	54.000	9.000	36.000	20.000	68.500
3 of 4	47.250	115.000	30.000	72.000	39.000	77.500	20.250	54.000	31.000	96.000

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

SYSTAT

程序代写代做 CS编程辅导



- SYSTAT is a Windows-based statistics platform. You can download a free version of SYSTAT from the link below.

- <https://systatsoftware.com/> WeChat: cstutorcs

Assignment Project Exam Help

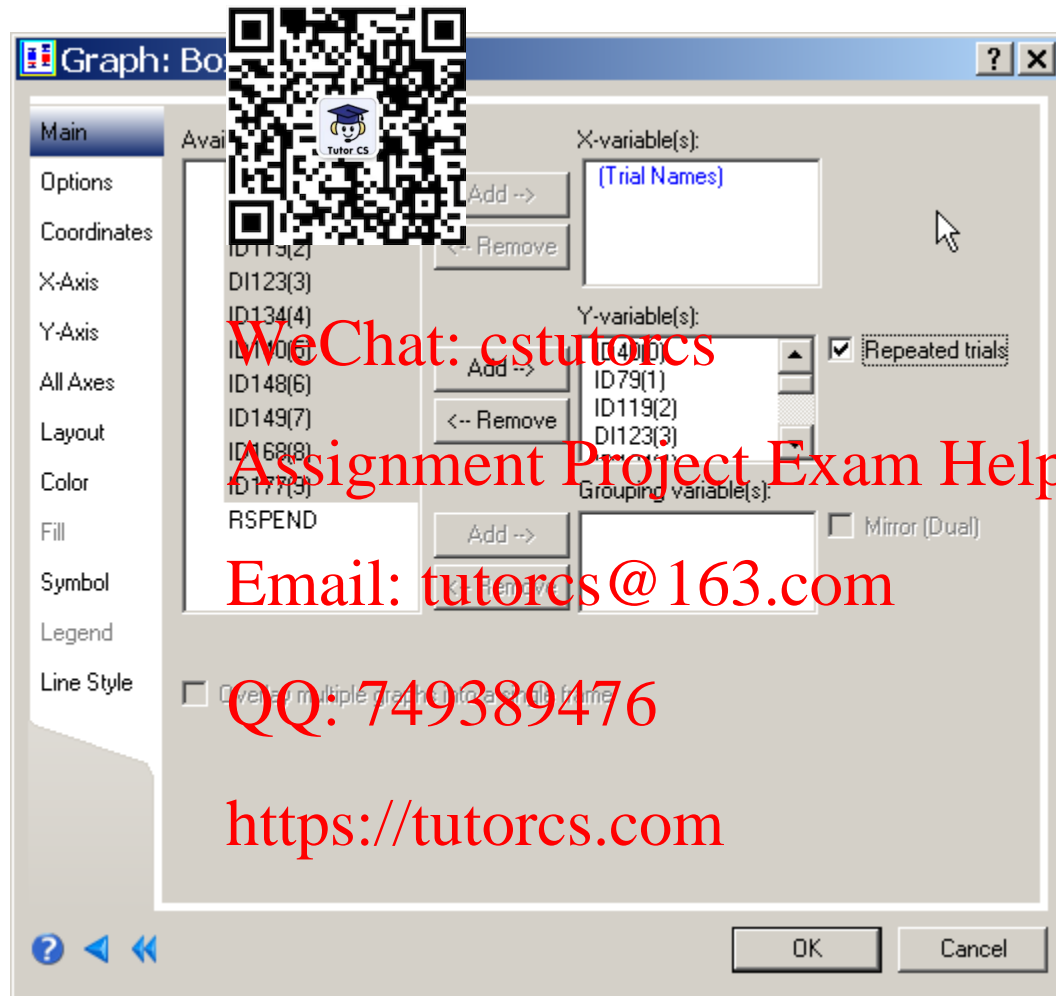
- <https://systatsoftware.com/products/systat/mystat-statistical-analysis-product-for-student-use/>

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Screenshot from SYSTAT... 程序代写代做 CS编程辅导



程序代写代做 CS编程辅导

Making sense of the data...



- How do we make sense of these information?
- What can we infer from
 - Descriptive statistics of the data
 - The distribution from stem and leaf plot
 - The box plot, etc...

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

<https://flux.qa> ^{程序代写 代做 CS 编程辅导} (Feed code: SJ6KGV)

Question 4



From the boxplot, which customer has the greater median amount spent?

A. ID79

B. ID123

C. ID140

D. ID177

E. None of the above.

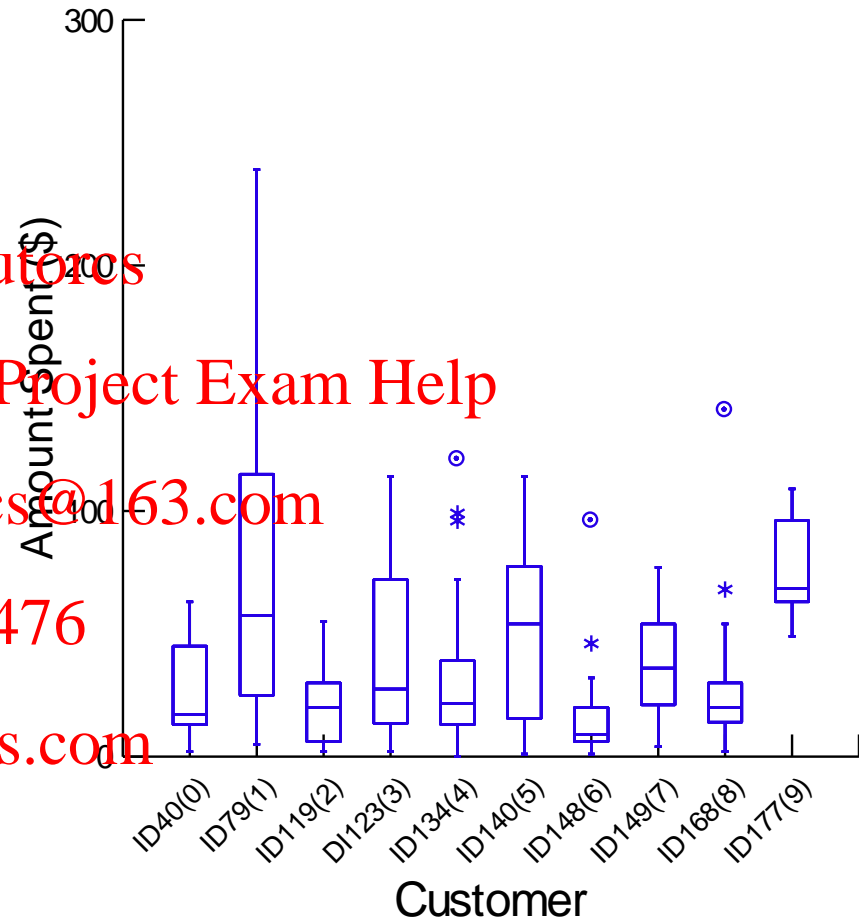
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



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

Question 5



From the boxplot, the most “inconsistent” customer?

A. ID79

B. ID123

C. ID140

D. ID177

E. None of the above.

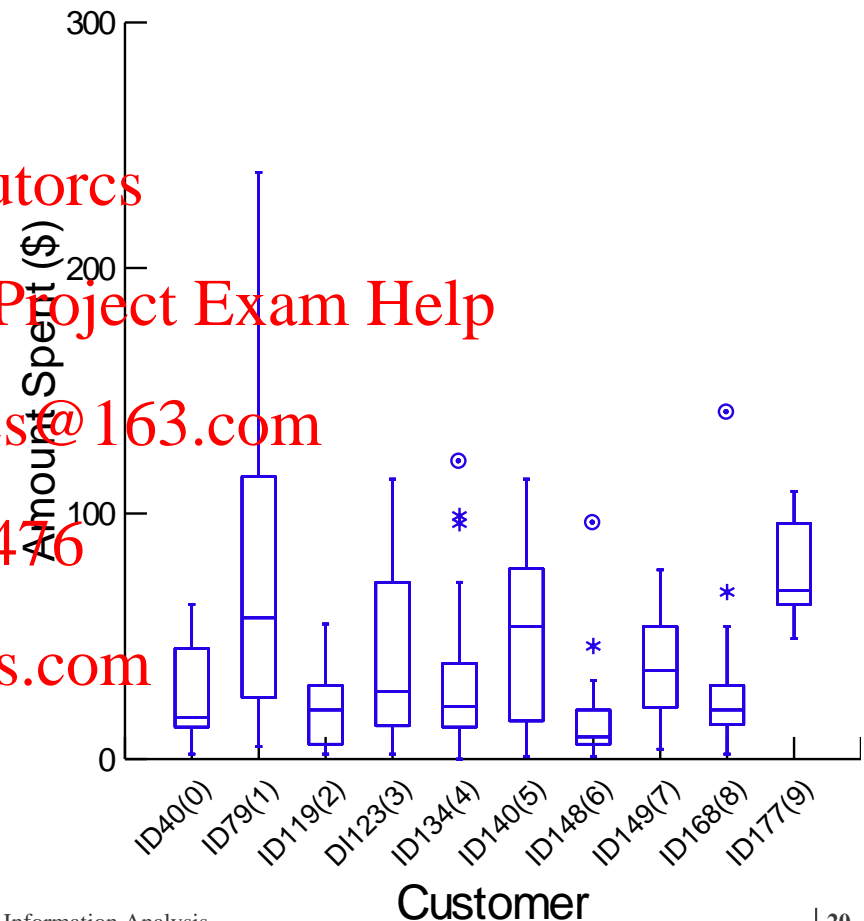
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



<https://flux.qa> 程序代写 代做 CS 编程辅导
(Feed code: SJ6KGV)

Question 6



From the boxplot, which is the best customer?

A. ID79

B. ID123

C. ID140

D. ID177

E. None of the above.

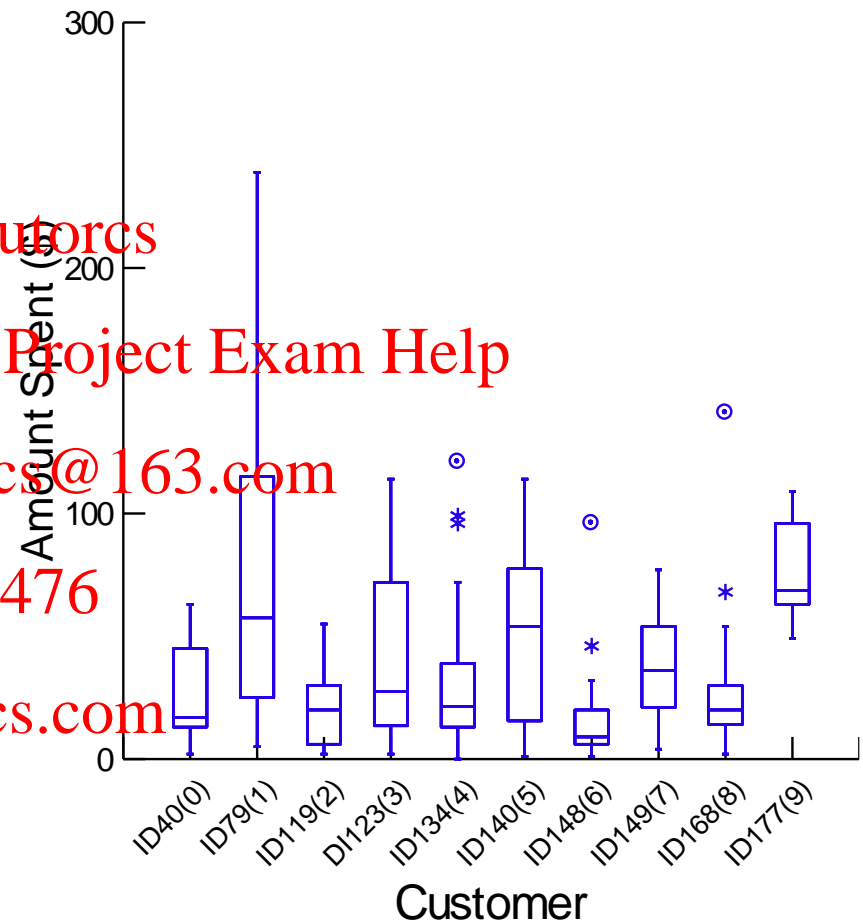
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com



程序代写代做 CS编程辅导

Measures of spread

- The variance is the average of the squared deviations about the mean for estimation of the mean.
- The standard deviation is the most well known. It is the square root of the variance.
- The range is largest – smallest observation.
- The interquartile range is $Q3 - Q1$ it contains the middle 50% of observations.

WeChat: estutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

Let's have a look at the 'shape' of distribution...

<https://flux.qa> ^{程序代写代做CS编程辅导} (Feed code: SJ6KGV)

Question 7

The histogram corresponds to which data set:



A. 1

B. 2

C. 3

D. 4

E. 5

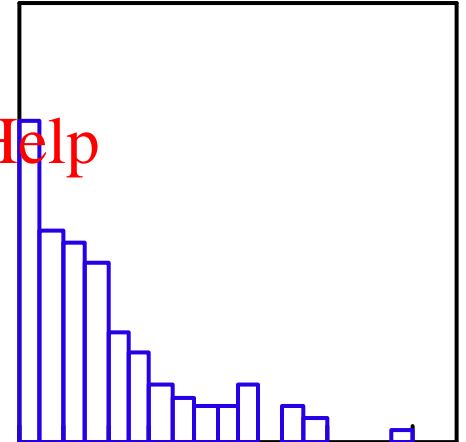
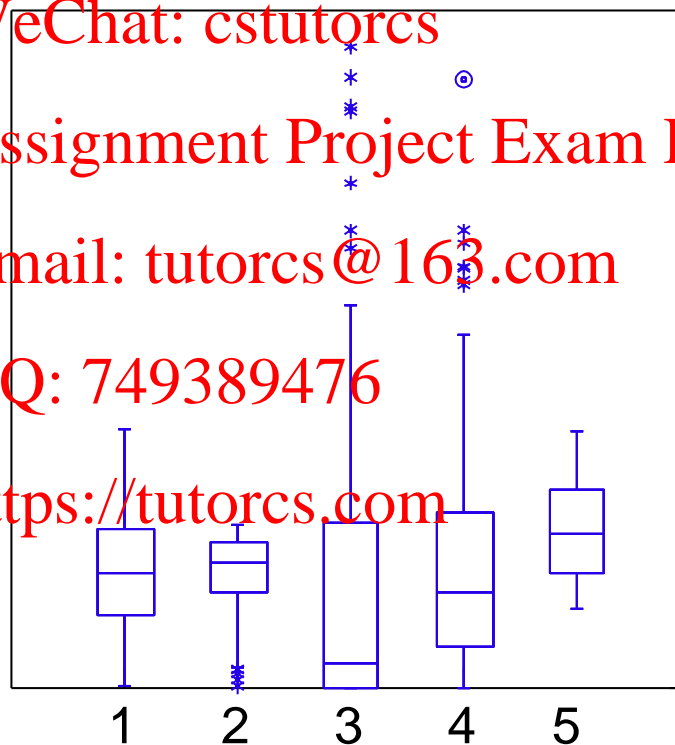
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



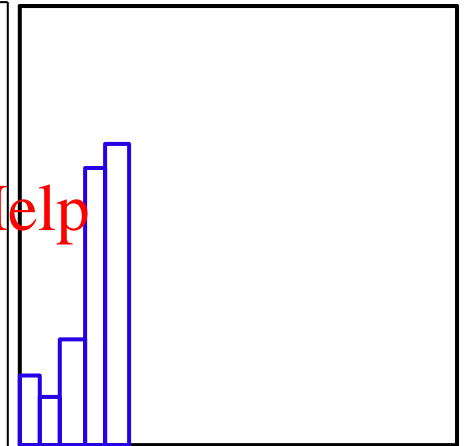
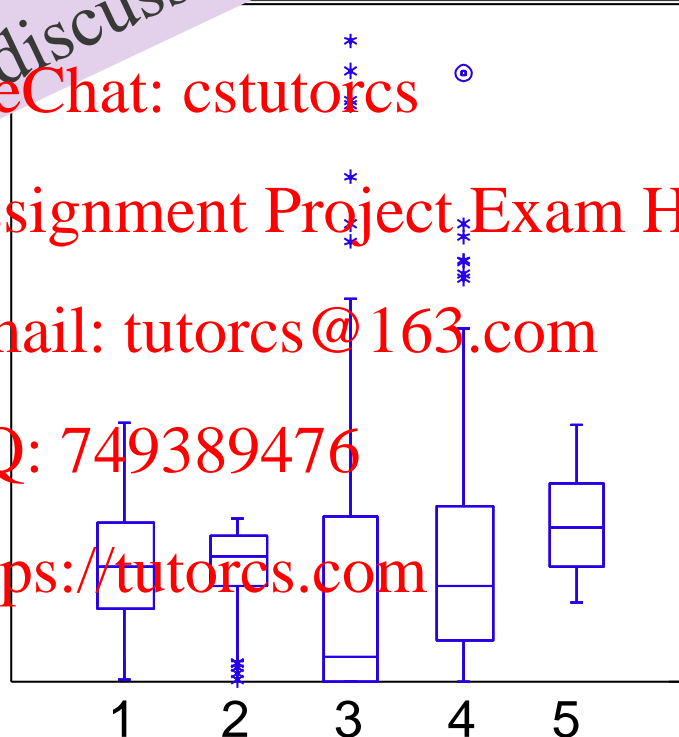
<https://flux.qa> 程序代码代做CS编程辅导

Question 8

The histogram for the variable 'Area' corresponds to which data set:



- A. 1
- B. 2
- C. 3
- D. 4
- E. 5



<https://flux.qa>

程序代写代做CS编程辅导
(Feed code: SJ6K7)

Question 9



The histogram for the variable 'error' corresponds to which data set:

- A. 1
- B. 2
- C. 3
- D. 4
- E. 5

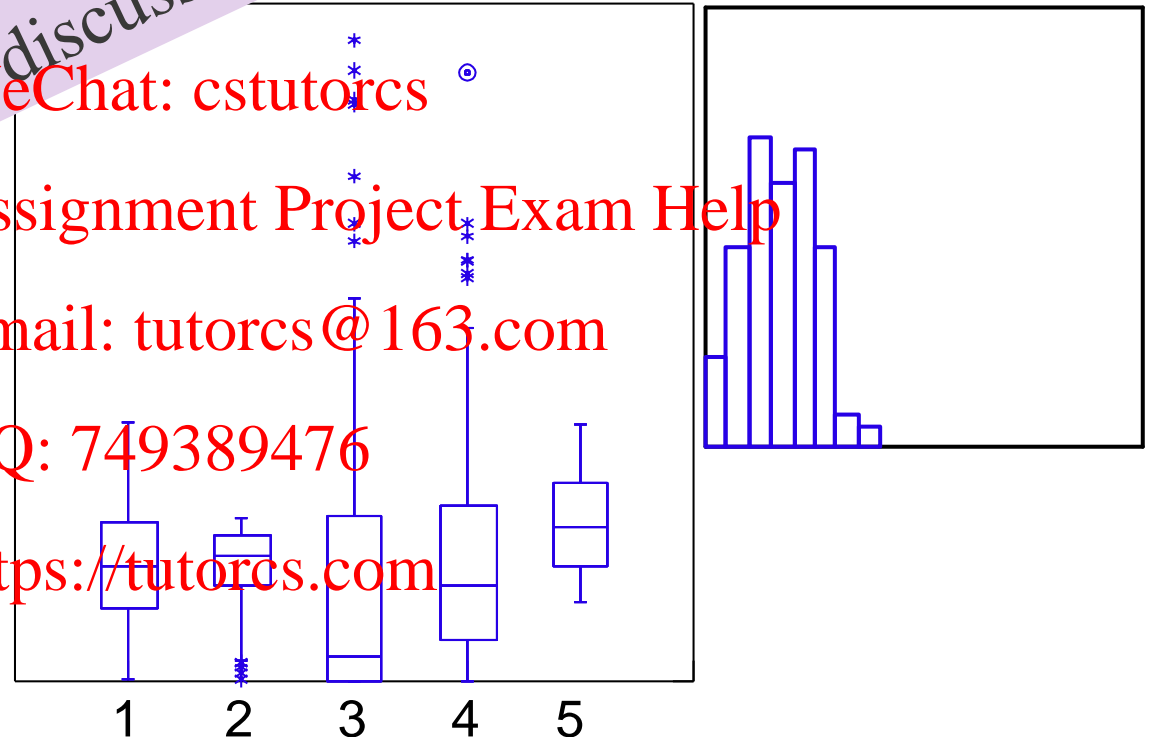
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

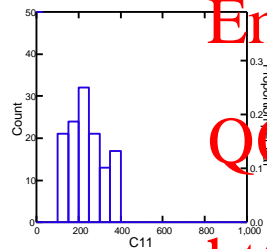
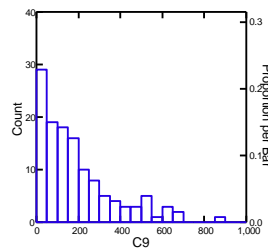
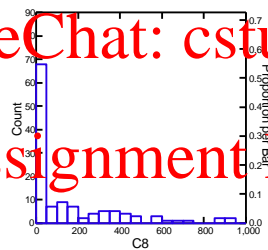
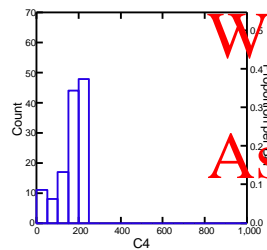
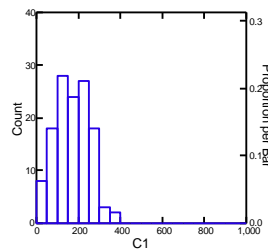


Distribution Shape and Boxplot

程序代写代做 CS编程辅导



- Here's 5 distributions and the corresponding boxplots



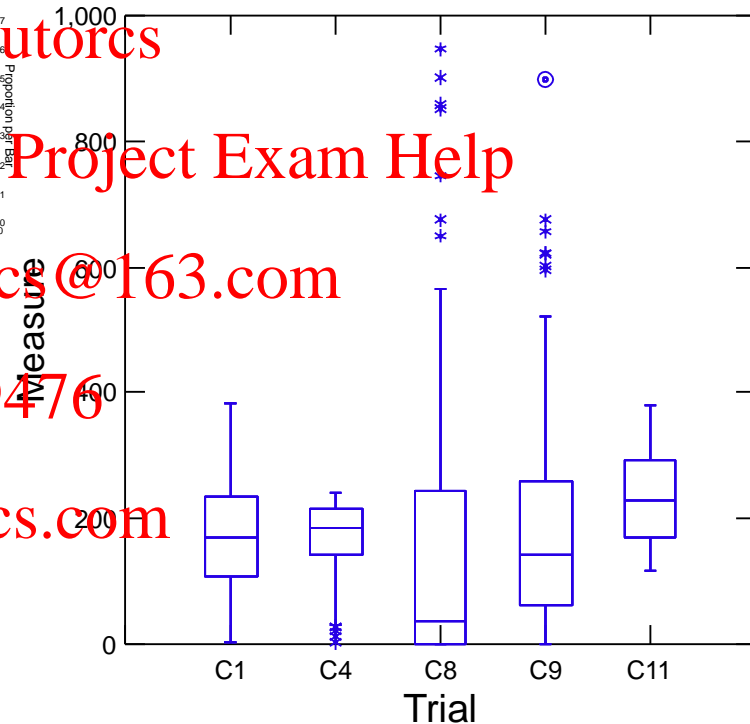
WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

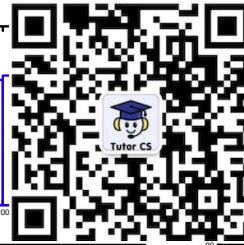
QQ: 749389476

<https://tutorcs.com>



Full Set

程序代写代做 CS编程辅导



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 calculate the basic descriptive statistics using Excel and SYSTAT;
- Plot histograms and boxplots of data, including several groups of data on a single plot using SYSTAT



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



- Reading: Graphical / Numerical Descriptive Methods
 - 7th Ed. Sections 2.1, 3.1, 4.1, 4.4, 5.1 - 5.3.

WeChat: cstutorcs

- Questions: Graphical / Numerical Descriptive Methods
 - 7th Ed. 5.17, 5.41, 5.45, 5.46, 5.67, 5.70.

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

- Tutorial 3 Questions

<https://tutorcs.com>