EECE5640 - SIMULATION AND PERFORMANCE EVALUATION: HOMEWORK 4

ANNA DEVRIES[1]

25 March 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

[1] *Department of Electrical and Computer Engineering, Northeastern University, Boston, United States*

1 QUESTION 1

Description: Ex. 4.1.11: Calculate $\bar{x}$ and s by hand, using the two-pass algorithm, the one pass algorithm, and Welford's algorithm in the following two cases.

- Data based on n = 3 observations: $x_1 = 1$, $x_2 = 6$, $x_3 = 2$.
- The sample path $x(t) = 3$ for $0 < t \leqslant 2$, and $x(t) = 8$ for $2 < t \leqslant 5$, over the time interval $0 < t < 5$.

1.1 First Data Set Solution

1.1.1 *Two-Pass Algorithm with First Data Set*

The sample standard deviation equation is $s = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$. Now this requires two passes through the data, first to compute the mean $\bar{x}$ and second to compute the squared differences about $\bar{x}$. Additionally, the data must all be stored (or re-created) in order to pass it twice.
First pass through:

sample 1: $x_1 = 1$ $\quad \Sigma x_i = 0 + 1 = 1$
sample 2: $x_2 = 6$ $\quad \Sigma x_i = 1 + 6 = 7$
sample 3: $x_3 = 2$ $\quad \Sigma x_i = 7 + 2 = 9$
therefore $\bar{x} = \frac{\Sigma x_i}{n} = \frac{9}{3} = 3$

Second pass through:

sample 1: $x_1 = 1$ $\quad \Sigma(x_i - \bar{x})^2 = 0 + (1-3)^2 = 4$
sample 2: $x_2 = 6$ $\quad \Sigma(x_i - \bar{x})^2 = 4 + (6-3)^2 = 13$
sample 3: $x_3 = 2$ $\quad \Sigma(x_i - \bar{x})^2 = 13 + (2-3)^2 = 14$
therefore $s = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2} = \sqrt{\frac{1}{n}(14)} = 2.16$

1.1.2 *One-Pass Algorithm with First Data Set*

The sample standard deviation, however, can be simplified into a one-pass method. This method requires one pass through the data set, summing the squared value of x and summing the sample values, and calculating standard deviation and mean at the end.

sample 1: $x_1 = 1$ $\quad \Sigma x_i^2 = 0 + 1^2 = 1$ $\quad \Sigma x_i = 0 + 1 = 1$
sample 2: $x_2 = 6$ $\quad \Sigma x_i^2 = 1 + 6^2 = 37$ $\quad \Sigma x_i = 1 + 6 = 7$
sample 3: $x_3 = 2$ $\quad \Sigma x_i^2 = 37 + 2^2 = 41$ $\quad \Sigma x_i = 7 + 2 = 9$
therefore $\bar{x} = \frac{\Sigma x_i}{n} = \frac{9}{3} = 3$ and $s = \sqrt{(\frac{1}{n}\Sigma x_i^2) - \bar{x}^2} = \sqrt{(\frac{1}{3} * 41) - 3^2} = \sqrt{4.67} = 2.16$

1.1.3 *Welford's One-Pass Algorithm with First Data Set*

This method is less prone to accumulated round-off error that can be found in the conventional one-pass algorithm. It works by keeping a running sample mean and sample sum of squared deviations. The $\bar{x_i}$ and v_i are computed recursively. No prior knowledge of the sample size n is required, rather it updates the values as we go through the sample set.

sample 1: $x_1 = 1$
$n = 0 + 1 = 1$ $\quad d = x_i - \bar{x} = 1 - 0 = 1$
$v = v + d * d * (n-1)/n = 0 + 1 * 1 * (1-1)/1 = 0$ $\quad \bar{x} = \bar{x} + d/n = 0 + 1/1 = 1$
sample 2: $x_2 = 6$
$n = 1 + 1 = 2$ $\quad d = x_i - \bar{x} = 6 - 1 = 5$
$v = v + d * d * (n-1)/n = 0 + 5 * 5 * (2-1)/2 = 12.5$ $\quad \bar{x} = \bar{x} + d/n = 1 + 5/2 = 3.5$
sample 3: $x_3 = 2$
$n = 2 + 1 = 3$ $\quad d = x_i - \bar{x} = 2 - 3.5 = -1.5$

$v = v + d * d * (n-1)/n = 12.5 + (-1.5) * (-1.5) * (3-1)/3 = 14$
$\bar{x} = \bar{x} + d/n = 3.5 + (-1.5/3) = 3$
therefore, $\bar{x} = 3$ and $s = \sqrt{\frac{v}{n}} = \sqrt{\frac{14}{3}} = 2.16$

1.2 Second Data Set Solution

1.2.1 *Two-Pass Algorithm with Second Data Set*

Since the second data set is continuous and not discrete, I utilized the following formulas to calculate the mean and standard deviation: $\bar{x} = \frac{1}{\tau}\int_0^\tau x(t)dt = \frac{1}{t_n}\sum_{i=1}^{n} x_i\delta_i$ and $s^2 = \frac{1}{\tau}\int_0^\tau (x(t)-\bar{x})^2 dt = \frac{1}{t_n}\sum_{i=1}^{n}(x_i - \bar{x})^2\delta_i$ such that $\delta_i = (t_i - t_{i-1})$.
This algorithm requires two-passes on the sample space - first to acquire the mean and second to calculate the standard deviation.
First Pass Through:
sample one: $x_1 = 3$ for $(0, 2]$
$\bar{x}_1 = \bar{x}_{i-1} + \sum_{i=1}^{n} x_i\delta_i = 0 + (3 * (2-0)) = 6$
sample two: $x_2 = 8$ for $(2, 5]$
$\bar{x}_2 = \bar{x}_{i-1} + \sum_{i=1}^{n} x_i\delta_i = 6 + (8 * (5-2)) = 30$
therefore $\bar{x} = \frac{1}{5} * 30 = 6$
Second Pass Through:
sample one: $x_1 = 3$ for $(0, 2]$
$s^2 = \frac{1}{\tau}\int_0^\tau (x(t)-\bar{x})^2 dt = \frac{1}{t_n}\sum_{i=1}^{n}(x_i - \bar{x})^2\delta_i$
$s^2 = ((3-6)^2 * (2-0)) + 0 = (9 * (2-0)) + 0 = 18$
sample two: $x_2 = 8$ for $(2, 5]$
$s^2 = 18 + ((8-6)^2 * (5-2)) = 30$
therefore $s = \sqrt{\frac{1}{5} * 30} = 2.45$

1.2.2 *One-Pass Algorithm with Second Data Set*

This requires only one pass through the continuous data set. I utilized the following formulas for mean and standard deviation: $\bar{x} = \frac{1}{\tau}\int_0^\tau x(t)dt = \frac{1}{t_n}\sum_{i=1}^{n} x_i\delta_i$ and $s^2 = \frac{1}{\tau}\int_0^\tau (x(t)-\bar{x})^2 dt = \frac{1}{t_n}\sum_{i=1}^{n}(x_i - \bar{x})^2\delta_i$ = $(\frac{1}{t_n}\sum_{i=1}^{n}(x_i^2\delta_i)) - \bar{x}^2$ such that $\delta_i = (t_i - t_{i-1})$.
sample one: $x_1 = 3$ for $(0, 2]$
$\bar{x} = 0 + (3 * (2-0)) = 6$
$s^2 = 0 + (3^2 * (2-0)) = 18$
sample two: $x_2 = 8$ for $(2, 5]$
$\bar{x} = 6 + (8 * (5-2)) = 30$
$s^2 = 18 + (8^2 * (5-2)) = 210$
therefore $\bar{x} = \frac{1}{5} * 30 = 6$ and $s = \sqrt{\frac{1}{5} * 210 - 6^2} = 2.45$

1.2.3 *Welford's One-Pass Algorithm with Second Data Set*

This requires only one pass through the continuous data set. I utilized the following formulas: $\bar{x}_i = \bar{x}_{i-1} + \frac{\delta_i}{t_i}(x_i - \bar{x}_{i-1})$ and $v_i = v_{i-1} + \frac{\delta_i * t_{i-1}}{t_i}(x_i - \bar{x}_{i-1})^2$.
sample one: $x_1 = 3$ for $(0, 2]$
$\bar{x}_1 = 0 + \frac{2-0}{2}(3-0) = 3$
$v_1 = 0 + \frac{(2-0)*0}{2}(3-0)^2 = 0$
sample two: $x_2 = 8$ for $(2, 5]$
$\bar{x}_2 = 3 + \frac{5-2}{5}(8-3) = 6$

$v_2 = 0 + \frac{(5-2)*2}{5}(8-3)^2 = 30$
therefore $\bar{x} = \bar{x}_2 = 6$ and $s = \sqrt{\frac{v_n}{t_n}} = \sqrt{\frac{30}{5}} = \sqrt{6} = 2.45$

2 QUESTION 2

Description: Ex. 4.2.2

- Generate the 2000-ball histogram in Example 4.2.2.
- Generate the corresponding histogram if 10,000 balls are placed, at random, in 1000 boxes.
- Calculate the histogram mean ($\bar{x}$) and the histogram standard deviation (s) for both the 2000 balls and the 10,000 balls.

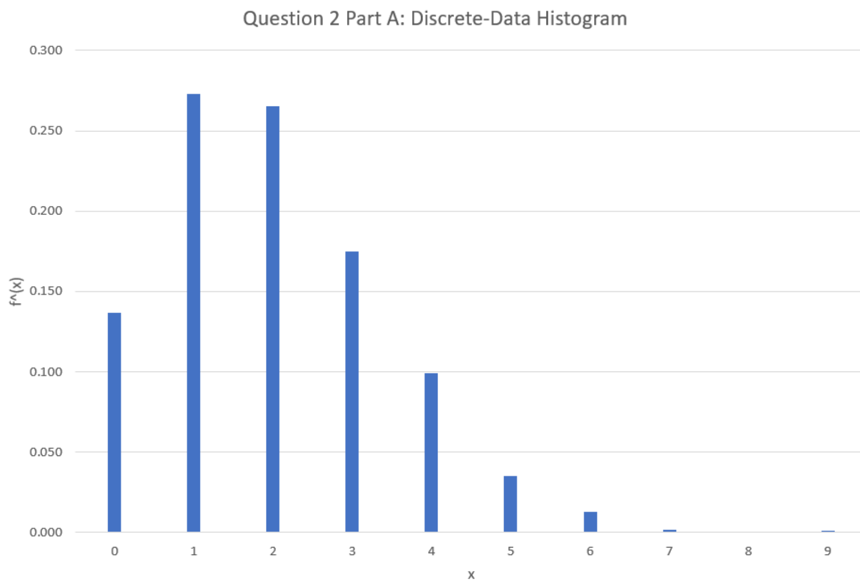
2.1 Solution

The results are in Figure 1. As shown, the 2,000 ball example gives a mean of 2 and standard deviation of 1.419; whereas, the 10,000 ball example gives a mean of 10 and standard deviation of 3.209. The histogram plots are shown in Figure 2.

```
z3r0@ubuntu:~/Desktop/eecec5640$ gcc hw4_2a.c -lm rng.o -o hw4 && ./hw4

Histogram Problem
Author: Anna DeVries
Date: 17 March 2021
Description: 2000 balls are randomly thrown into 1000 containers
Output: histogram mean and standard deviation
--------------------------------------------------------------------------------------

Histogram {x, f^(x)}:           {0, 0.137}      {1, 0.273}      {2, 0.265}      {3, 0.175}      {4, 0.099}
                                {5, 0.035}      {6, 0.013}      {7, 0.002}      {8, 0.000}      {9, 0.001}

Histogram Mean :                2.00
Histogram Standard Deviation :  1.419
```

```
z3r0@ubuntu:~/Desktop/eecec5640$ gcc hw4_2b.c -lm rng.o -o hw4 && ./hw4

Histogram Problem
Author: Anna DeVries
Date: 17 March 2021
Description: 10000 balls are randomly thrown into 1000 containers
Output: histogram mean and standard deviation
--------------------------------------------------------------------------------------

Histogram {x, f^(x)}:           {0, 0.001}      {1, 0.000}      {2, 0.001}      {3, 0.004}      {4, 0.024}
                                {5, 0.031}      {6, 0.080}      {7, 0.085}      {8, 0.117}      {9, 0.122}
                                {10, 0.128}     {11, 0.109}     {12, 0.077}     {13, 0.071}     {14, 0.065}
                                {15, 0.032}     {16, 0.026}     {17, 0.010}     {18, 0.009}     {19, 0.004}
                                {20, 0.002}     {21, 0.002}
Histogram Mean :                10.00
Histogram Standard Deviation :  3.209
```

Figure 1: Question 2 Program Outputs for a) part a. b) part b.



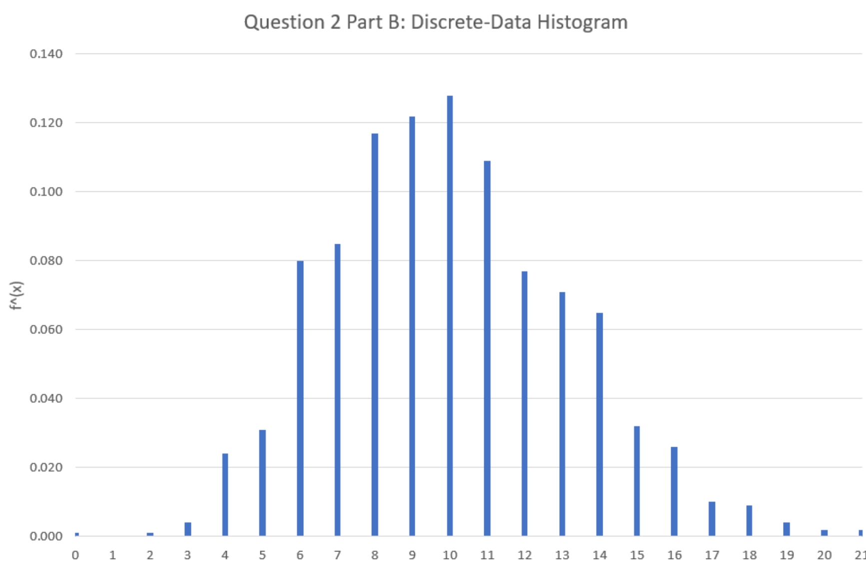


Figure 2: Question 2 Histograms for a) part a. b) part b.

3 QUESTION 3

Description: Ex. 4.2.11: A test is compiled by selecting 12 different questions, at random and without replacement, from a well-publicized list of 120 questions. After studying this list you are able to classify all 120 questions into two classes, I and II. Class I questions are those about which you feel confident; the remaining questions define class II. Assume that your grade probability, conditioned on the class of the problem, is

	A	B	C	D	F
class I	0.6	0.3	0.1	0.0	0.0
class II	0.0	0.1	0.4	0.4	0.1

Each test question is grade on an A = 4, B = 3, C = 2, D = 1, F = 0 scale and a score of 36 or better is required to pass the test.

- If there are 90 class-I questions in the list, use Monte Carlo simulation and 100000 replications to generate a discrete-data histogram of scores.
- From this histogram, what is the probability that you will pass the test?

3.1 Solution

The results are in Figure 3. As shown, the probability of passing the test is 56.74%. The histogram plot is shown in Figure 4.

```
z3r0@ubuntu:~/Desktop/eecec5640$ gcc hw4_3.c -lm rng.o -o hw4 && ./hw4

Histogram Problem
Author: Anna DeVries
Date: 18 March 2021
Description: utilizes a Monte Carlo simulation to generate a discrete-data histogram of your grade probability
Output: histogram and probability you will pass the test
-------------------------------------------------------------------------------------------------------------

Results after 100000 iterations:
        ...histogram {x:f^(x)}:         {17:0.00001}    {18:0.00001}    {19:0.00005}    {20:0.00014}    {21:0.00014}    {22:0.00024}    {23:0.00069}    {24:0.00146}
                                        {25:0.00299}    {26:0.00476}    {27:0.00798}    {28:0.01355}    {29:0.02012}    {30:0.02985}    {31:0.04306}    {32:0.05694}
                                        {33:0.07130}    {34:0.08230}    {35:0.09697}    {36:0.10170}    {37:0.10157}    {38:0.09479}    {39:0.08277}    {40:0.06589}
                                        {41:0.04969}    {42:0.03300}    {43:0.01977}    {44:0.01108}    {45:0.00483}    {46:0.00176}    {47:0.00049}    {48:0.00010}

        ...prob. of passing test:       56.74%

z3r0@ubuntu:~/Desktop/eecec5640$
```

Figure 3: Question 3 Program Output.

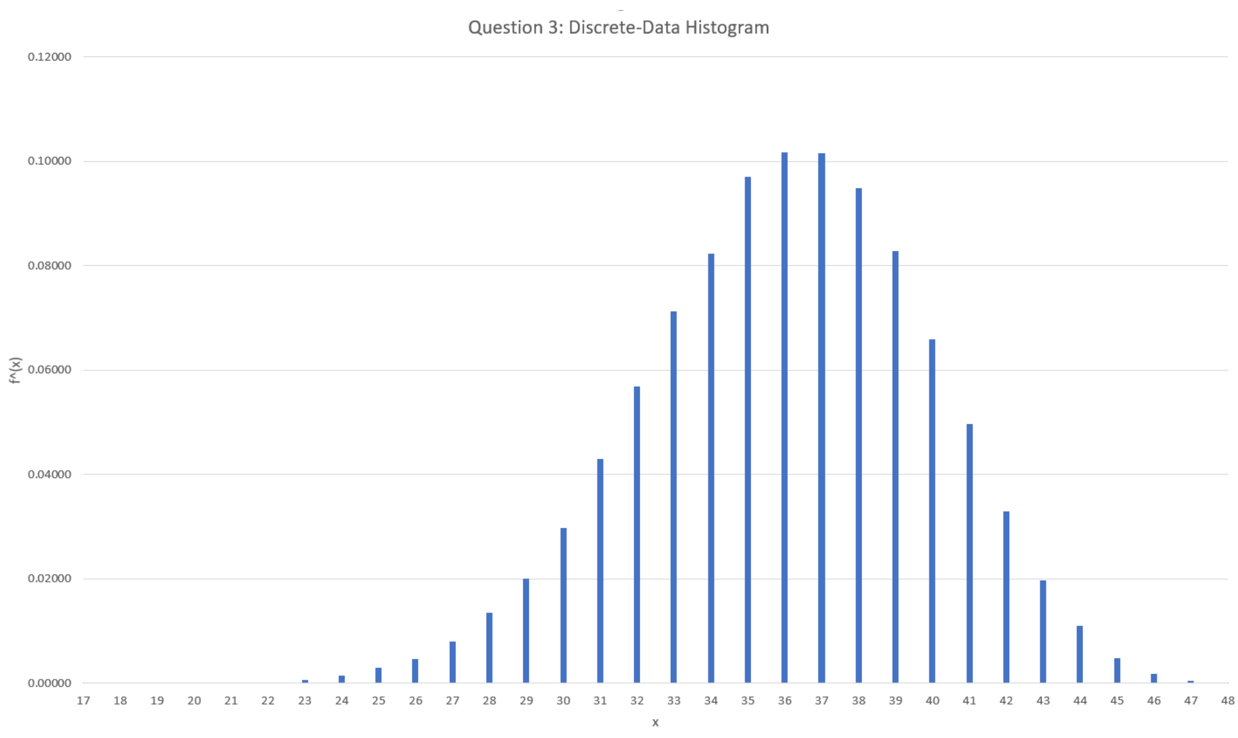


Figure 4: Question 3 Histogram.

4 QUESTION 4

Description: Ex. 4.3.5

- Construct a continuous-data histogram of the service times (in ac.dat - Ex. 1.2.6).
- Compare the histogram mean and standard deviation with the corresponding sample mean and standard deviation, and justify your choice of the histogram parameters a, b and either k or δ.

4.1 Solution

The results are in Figure 5.

```
z3r0@ubuntu:~/Desktop/eecec5640/hw4$ gcc hw4_4.c -lm rng.o -o hw4 && ./hw4

Continous Data Histogram Problem
Author: Anna DeVries
Date: 22 March 2021
Description:    Given a file ac.dat of <arrival_time departure_time>.
                This program converts data to service time assuming single-server queue.
                For the continous data histogram: a =   0, b =  16, k =  18.
Output: continous-data histogram, and histogram mean and standard deviation
---------------------------------------------------------------------------------------


Continious-Data Histogram:

          bin     midpoint     count   proportion    density
            1       0.444        33      0.066        0.074
            2       1.333        95      0.190        0.214
            3       2.222       120      0.240        0.270
            4       3.111       100      0.200        0.225
            5       4.000        51      0.102        0.115
            6       4.889        50      0.100        0.113
            7       5.778        25      0.050        0.056
            8       6.667        12      0.024        0.027
            9       7.556         6      0.012        0.014
           10       8.444         5      0.010        0.011
           11       9.333         1      0.002        0.002
           12      10.222         0      0.000        0.000
           13      11.111         1      0.002        0.002
           14      12.000         0      0.000        0.000
           15      12.889         0      0.000        0.000
           16      13.778         0      0.000        0.000
           17      14.667         0      0.000        0.000
           18      15.556         1      0.002        0.002

        sample size .... =     500
        mean ........... =   3.031
        stdev .......... =   1.829

---------------------------------------------------------------------------------------

Single Server Queue Information:

        sample size .... =     500
        mean ........... =   3.032
        stdev .......... =   1.823
        utilization .... =   0.740

        NOTE: this is sample information, it will differ from continious-data histogram.

z3r0@ubuntu:~/Desktop/eecec5640/hw4$ 
```

Figure 5: Question 4 Program Output.

For this question, I chose a = 0.0, b = 16.0, and k = 18. I explicitly chose a and b such that I minimized the amount of outliers. If I increased a by 1.0, the program produced 38 low outliers; however, if I decreased b by 1.0, the program produced 1 high outlier. After this, I sought to find a k value that kept the mean and standard deviation near the sample versions of the values. I obtained a range of possible k values as $\log_2(n) < k < \sqrt{n}$ with a bias towards $\frac{5}{3}\sqrt[3]{n}$. This produced a range of $8.97 < k < 22.36$ with a bias towards 13.22. With this in mind, I tried values within this range until the mean and standard deviation were most similar to the sample mean and sample standard deviation. I found this value to be 18. As shown the histogram mean and standard deviation are 3.031 and 1.829 respectively. The sample mean and standard deviation are 3.032 and 1.823 respectively. The values are slightly different due to Quantization error. This is associated with binning of continuous data; however, this error is very low (0.002 difference between the means and 0.006 difference between the standard deviations).

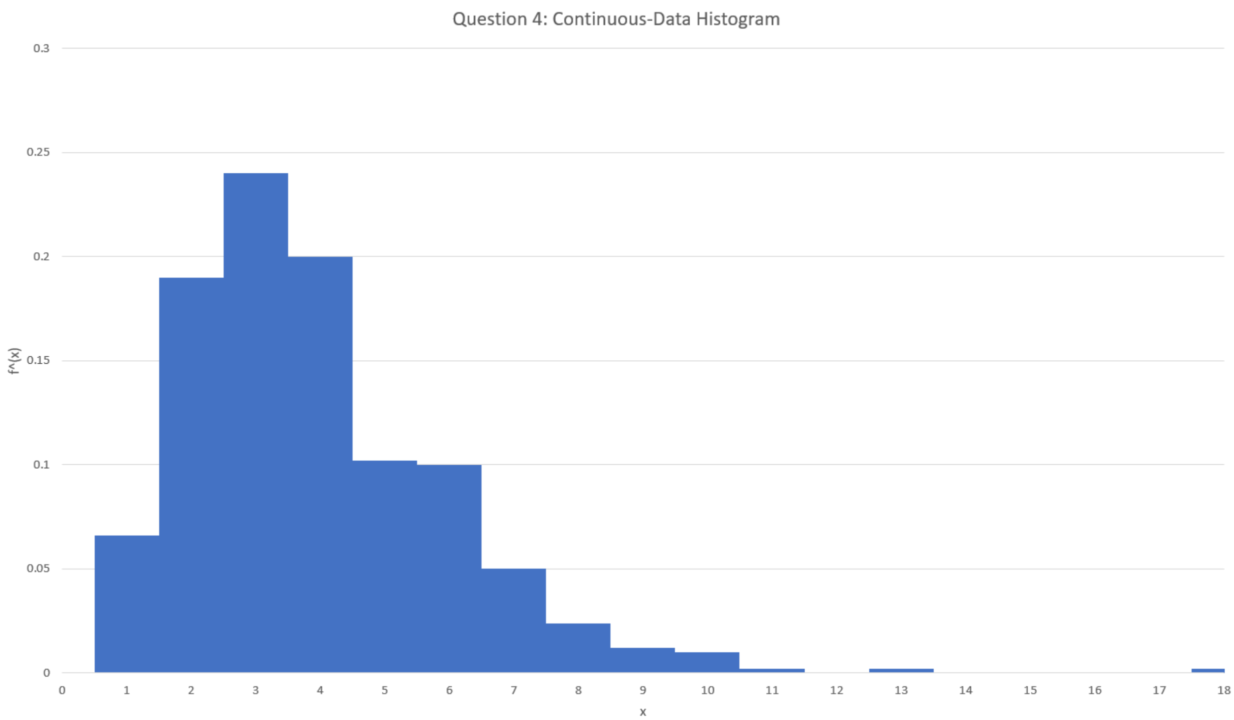


Figure 6: Question 4 Histogram.

The histogram plot is shown in Figure 6.

REFERENCES