

MAT 121 Statistics I

Midterm Exam #1

09/27/2023

Time: 90 minutes

Problem 1.

The following is a sample of ages (in months) of 18 children at a daycare:
18 19 22 22 24 24 25 26 28 29 29 30 31 32 35 36 36 42
What is the median age of the sample?

Answers *

☐ 29

☐ 28.2

☐ 30.5

☒ 28.5

☐ 31

IntroStatsApps: Descriptive Statistics

Types of Descriptive Statistics


Numerical Measures

comma separated numerical data

18, 19, 22, 22, 24, 24, 25, 26, 28, 29, 29, 30, 31, 32, 35, 36, 36, 42

Measure Types

measures of center



[Report bugs to C. Peng](#)

Measures of Center

The data values are:

18, 19, 22, 22, 24, 24, 25, 26, 28, 29, 29, 30, 31, 32, 35, 36, 36, 42

The sorted data values are:

18, 19, 22, 22, 24, 24, 25, 26, 28, 29, 29, 30, 31, 32, 35, 36, 36, 42

1. Sample (population) mean

$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = 28.2$, and $\mu = \sum_{i=1}^n \frac{x_i}{n} = 28.2$ (if this data set is a population).

2. Median

The median of a given data set is the middle number of **sorted** data set. Based on this definition, the median of the given data set is: 28.5

3. Mode

A data value that appears most frequently (frequency > 1) is called the mode of the data. Based on this definition, a set of data may have one mode, more than one mode, or no mode at all. Using the above definition, this data set has 4 modes: 22, 24, 29, 36.

Problem 2.

The following grouped frequency table of the income, x , of 30 employees at a local small business (in \$1000s).

Income	[26, 28]	(28, 30]	(30, 32]	(32, 34]	(34, 36]
Frequency	2	11	8	5	4

The relative cumulative frequency of the $28 < x \leq 30$ class is

Answers *

☐

11

☐

≈ 0.43

☐

≈ 0.06

☒

≈ 0.37

☐

≈ 0.7

$$\frac{11}{2+11+8+5+4} = \frac{11}{30} \approx 0.37.$$

Problem 3.

A study of 1106 college students asked about their preference for online resources. The following relative frequency distribution was determined based on the survey.

Resource	Relative Frequency
Google or Google Scholar	0.736
Library database or website	0.136
Wikipedia or online encyclopedia	0.094
Other	0.034

Of the 1106 students who participated in the survey, approximately how many chose Google or Google Scholar?

Answers *

- ☐ 34
- ☐ 292
- ☐ 736
- ☒ 814
- ☐ 921

$$1106 \times 0.736 = 814.016$$

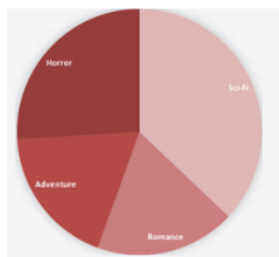
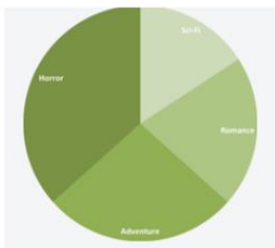
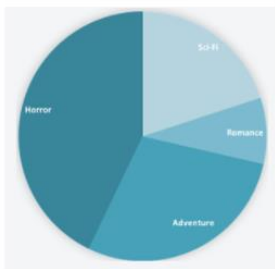
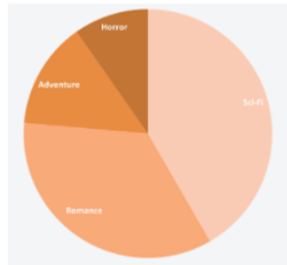
Problem 4.

The table below represents 360360 books grouped by their category:

Book category	Frequency
Science-fiction	150150
Romance	125125
Adventure	5050
Horror	3535

Draw a pie chart representing this table

Answers *



Matching numbers with the corresponding slices.

Problem 5

Find the mean, median, and mode for the following data set.

4 7 9 11 11 11 13 17 22 26

Answers *

☐

mode = 11, mean =12, median = 11

☐

mode = 11, mean =11, median = 11

☐

mode = 11, mean =13, median = 11.5

☒

mode = 11, mean =13, median = 11

☐


mode = 11, mean =14, median = 11

IntroStatsApps: Descriptive Statistics

Types of Descriptive Statistics
Numerical Measures

comma separated numerical data
4,7,9,11,11,11,13,17,22,26

Measure Types
measures of center


Report bugs to C. Peng

Measures of Center

The data values are:

4, 7, 9, 11, 11, 11, 13, 17, 22, 26

The sorted data values are:

4, 7, 9, 11, 11, 11, 13, 17, 22, 26

1. Sample (population) mean

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = 13.1, \text{ and } \mu = \sum_{i=1}^n \frac{x_i}{n} = 13.1 \text{ (if this data set is a population).}$$

2. Median

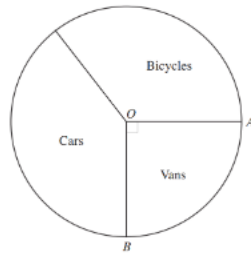
The median of a given data set is the middle number of **sorted** data set. Based on this definition, the median of the given data set is: 11

3. Mode

A data value that appears most frequently (frequency > 1) is called the mode of the data. Based on this definition, a set of data may have one mode, more than one mode, or no mode at all. Using the above definition, this data set has 1 mode: 11.

Problem 6.

The pie chart above, not drawn to scale, shows the number of vehicles parked outside a supermarket. Angle AOB is a right angle. Given that there were 60 vehicles, how many vans were there?



Answers *

☐

4

☐

6

☐

12

☒

15

☐

20

Based on the definition of the pie chart.

Problem 7

The mean temperature in Glens Falls for the month of February is 23 degrees with a standard deviation of 4.2 degrees. What is the z-score for a temperature of 17 degrees (keeping 3 decimal places)?

Answers *

☐

9.523

☒

-1.429

☐

1.429

☐

-2.928

☐

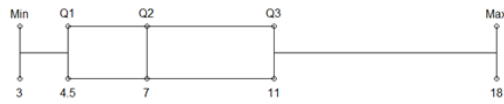
-0.340

$$z_{17} = \frac{23 - 17}{4.2} \approx 1.428571$$

Problem 8.

Construct a box-plot of data set: {3, 4, 4, 5, 5, 6, 8, 10, 10, 12, 15, 18}

Answers *




IntroStatsApps: Descriptive Statistics

Types of Descriptive Statistics
Numerical Measures

comma separated numerical data
3,4,4,5,5,6,8,10,10,12,15,18

Measure Types
5-number summary and boxplot



Report bugs to C. Peng

Five Number Summary and Boxplot

The data values are:
3, 4, 4, 5, 5, 6, 8, 10, 10, 12, 15, 18

1. Five Number Summary :

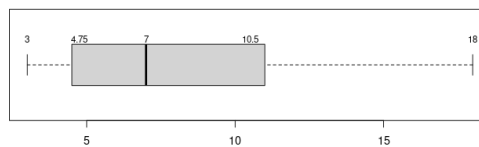
The five-number summary is used to describe the shape of the distribution of a given numerical data. It consists of five numbers: minimum data value, first quartile, median, the third quartile, and the maximum data value.

The five-number summary of this given data set is:

stats	value
Min.	3.00
1st Qu.	4.75
Median	7.00
3rd Qu.	10.50
Max.	18.00

2. Boxplot :

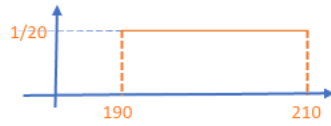
The boxplot is a geometric representation of the five-number summary. The boxplot of the given data set is given below.



Choose a one that is close to the true one.

Problem 9

The cholesterol content of large chicken eggs is **uniformly** distributed between 190 and 210 milligrams. The density curve has a rectangular shape.



What proportion of these eggs has cholesterol content above 205 milligrams?

Answers *



5/20



205/210



190/210



20/210

$$P(X > 205) = \frac{1}{20} (210 - 205) = \frac{1}{20} \times 5 = \frac{5}{20}$$

Problem 10

Find the variance of the following **sample** data set taken from a population.

4 11 11 13 15 18

Answers *

☐

112/6

☒

112/5

☐

72/6

☐

72/5

☐

976/5

IntroStatsApps: Descriptive Statistics

Types of Descriptive Statistics


Numerical Measures

comma separated numerical data

4,11,11,13,15,18

Measure Types

measures of variation



Report bugs to C. Peng

Measures of Variation

The data values are:
4, 11, 11, 13, 15, 18

The sorted data values are:
4, 11, 11, 13, 15, 18

1. Sample (population) variance

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} = 22.4, \text{ and } \sigma^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n} = 18.7 \text{ (if this data set is a population)}$$

2. Sample (population) standard deviation

The standard deviation is the square root of variance. Therefore, the both standard deviations are: 12

$$s = \sqrt{s^2} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}} = 4.7, \text{ and } \sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu)^2}{n}} = 4.3 \text{ (if this data set is a population)}$$

3. Inter - quartile range (IQR)

The inter-quartile range is defined to be the difference between the first and third quartiles. By the definition,
 $IQR = P_{75} - P_{25} = 15 - 11 = 4$.

$$\frac{112}{5} = 22.4$$

Problem 12

Fifteen percent of the students in a school of Business Administration are majoring in Economics, 20% in Finance, 35% in Management, and 30% in Accounting. The graphical device(s) which can be used to present these data is (are)

Answers *

☐

a line graph

☐

only a bar graph

☐

only a pie chart

☒

both a bar graph and a pie chart

☐

histogram

Problem 13

A scientist obtained a normally distributed population of scores with a mean of 70 and a standard deviation of 10. What proportion of scores do you expect to find in the interval between 60 to 80?

Answers *

- ☐ 1.00
- ☐ 0.50
- ☐ 0.34
- ☒ 0.68
- ☐ cannot be determined

1. What to Find?

- ☒ Probability (P_0)
- ☐ Percentile (X_0)

2. Which Probability?

- ☒ $P[V_0 < X < V_1] = ?$
- ☐ $P[X > V_0] = ?$
- ☐ $P[X < V_0] = ?$

Given Value #1: V_0

60

Given Value #2: V_1

80

3. Input Information

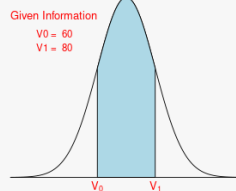
Population Mean: μ

70

Population Standard Deviation: σ

10

Sampling Distribution



Question:

$$P(60 < X < 80) = ?$$

Step 4. Finding P_0 :

$$\begin{aligned} P\left(\frac{60 - 70}{10} < \frac{X - 70}{10} < \frac{80 - 70}{10}\right) \\ &= P(Z_0 < Z < Z_1) \\ &= P(Z < Z_1) - P(Z < Z_0) \\ &= P(Z < 1) - P(Z < -1) \\ &= 0.8413 - 0.1587 = 0.6826 \end{aligned}$$

Step 5. Answer:

$$P(60 < X < 80) = 0.6826$$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 70}{10}$$

Step 2. Z-scores for V_0 and V_1

$$Z_0 = \frac{60 - 70}{10} = -1$$

$$Z_1 = \frac{80 - 70}{10} = 1$$

Step 3. Finding the left-tail Probabilities

$$P(Z < 1) = 0.8413$$

$$P(Z < -1) = 0.1587$$

Problem 14

The scoring of modern IQ is such that Intelligence Quotients (IQs) have a normal distribution of $\mu = 100$ and $\sigma = 15$.

Mensa International is a non-profit organization that accepts only people with IQ scores within the top 2%. What level of IQ qualifies one to be a member of Mensa?

Answers *

☐

115

☒

130.8

☐

145

☐

120

☐

cannot be determined

1. What to Find?

- ☐ Probability (P_0)
☒ Percentile (X_0)

2. X_0 in Which Probability?

- ☐ $P[X_0 < X < V] = P_0$
☐ $P[V < X < X_0] = P_0$
☒ $P[X > X_0] = P_0$
☐ $P[X < X_0] = P_0$

Given Probability: P_0

0.02

3. Input Information

Population Mean: μ

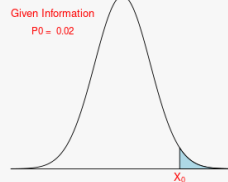
100

Population Standard Deviation: σ

15



Sampling Distribution



Question: Given $P(X > X_0) = 0.02$, what is X_0 ?

Step 3. Finding X_0 :

$$P\left(\frac{X - 100}{15} > \frac{X_0 - 100}{15}\right) = 0.02$$

or equivalently,

$$P\left(Z > \frac{X_0 - 100}{15}\right) = 0.02$$

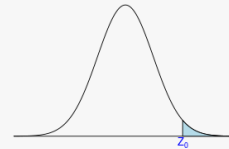
Therefore,

$$\frac{X_0 - 100}{15} = Z_0 = 2.05.$$

Step 4. Answer:

$$X_0 = 100 + (2.05) \times 15 = 130.75$$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 100}{15}.$$

Step 2. Find the Z-score corresponding to X_0

$$P(Z > Z_0) = 0.02,$$

or equivalent, $1 - P(Z < Z_0) = 0.02$.

Therefore, $Z_0 = 2.05$.

Problem 15.

A physical fitness association is including the mile run in its secondary-school fitness test. The time for this event for boys in secondary school is known to possess a normal distribution with a mean of 440 seconds and a standard deviation of 60 seconds. Find the probability that a randomly selected boy in secondary school can run the mile in less than 302 seconds.

Answers *

☐

0.9893

☒

0.0107

☐

0.5107

☐

0.4893

☐

cannot be determined

1. What to Find?

- ☒ Probability (P_0)
☐ Percentile (X_0)

2. Which Probability?

- ☐ $P[V_0 < X < V_1] = ?$
☐ $P[X > V_0] = ?$
☒ $P[X < V_0] = ?$

Given Value: V_0

302

3. Input Information

Population Mean: μ

440

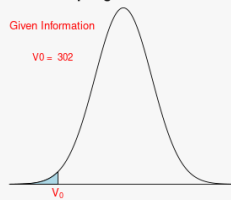
Population Standard Deviation: σ

60

Sampling Distribution

Given Information

$V_0 = 302$



Question:

$$P(X < 302) = ?$$

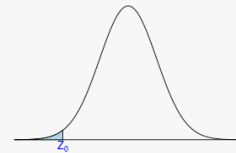
Step 4. Finding P_0 :

$$\begin{aligned} P\left(\frac{X - 440}{60} < \frac{302 - 440}{60}\right) \\ = P(Z < Z_0) \\ = P(Z < -2.3) = 0.0107 \end{aligned}$$

Step 5. Answer:

$$P(X < 302) = 0.0107$$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 440}{60}$$

Step 2. Z-scores for V

$$Z_0 = \frac{302 - 440}{60} = -2.3$$

Step 3. Finding the left-tail Probabilities

$$P(Z < -2.3) = 0.0107$$

Problem 16

The following are 40 measurements of the iron-solution index of tin-plate specimens, designed to measure the corrosion resistance of tin-plated steel. The original data set has been sorted in ascending order as:

16, 26, 28, 28, 28, 28, 30, 32, 34, 35, 36, 36, 37, 37, 40, 40, 40, 41, 41, 41,
42, 42, 42, 43, 43, 43, 44, 44, 44, 44, 45, 45, 45, 45, 45, 45, 46, 46, 46, 46,

We want to Construct a frequency table with **five** rows. Which of the following histograms is correct?

Answers *

☐

[15,31]	7
(31,42]	16
(42,53]	17
(53,64]	0
(64,75]	0

☐

[15,25]	1
(25,35]	9
(35,40]	7
(40,45]	19
(45,50]	4

☐

[15,23]	1
(23,31]	6
(31,39]	7
(39,46]	26
(46,53]	0

☒

[15,22]	1
(22,29]	5
(29,36]	6
(36,43]	14
(43,50]	14

Based on the procedure of constructing a histogram of a numerical data set.

Problem 17.

The age distribution of students at a community college is given below.

<u>Age (years)</u>	<u>Number of students (f)</u>
Under 21	2189
21-25	2031
26-30	1073
31-35	853
Over 35	221

A student from the community college is selected at random. The event E is defined as follows.

E = event the student is between 26 and 35 inclusive.

Determine the number of outcomes that comprise the event \bar{E} (i.e., not in E).

Answers *



4441



5294



4220



1926

$$2189 + 2031 + 221 = 4441.$$

Problem 18

Tomkins Associates reports that the mean clear height for a Class A warehouse in the United States is 22 feet. Suppose clear heights are normally distributed and that the standard deviation is 4 feet. A Class A warehouse in the United States is randomly selected,

What is the probability that the clear height is greater than 17 feet?

Answers *

☐

0.1056

☒

0.8994

☐

1.25

☐

-1.25

1. What to Find?

- ☒ Probability (P_0)
☐ Percentile (X_0)

2. Which Probability?

- ☐ $P[V_0 < X < V_1] = ?$
☒ $P[X > V_0] = ?$
☐ $P[X < V_0] = ?$

Given Value: V_0

17

3. Input Information

Population Mean: μ

22

Population Standard Deviation: σ

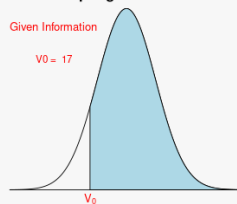
4

HESTER UNIV

Sampling Distribution

Given Information

$V_0 = 17$



Question:

$$P(X > 17) = ?$$

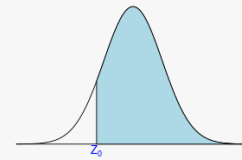
Step 4. Finding P_0 :

$$\begin{aligned} P\left(\frac{X - 22}{4} > \frac{17 - 22}{4}\right) \\ &= P(Z > Z_1) \\ &= 1 - P(Z < -1.25) \\ &= 1 - 0.1056 = 0.8944 \end{aligned}$$

Step 5. Answer:

$$P(X > 17) = 0.8944$$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 22}{4}$$

Step 2. Z-scores for V

$$Z_0 = \frac{17 - 22}{4} = -1.25$$

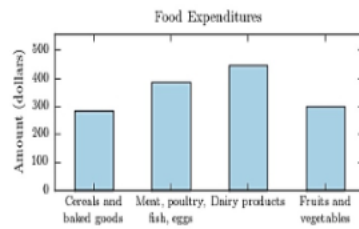
Step 3. Finding the left-tail Probabilities

$$P(Z < -1.25) = 0.1056$$

Problem 19

The following bar graph presents the average amount a certain family spent, in dollars, on various food categories in a recent year.

On which food category was the most money spent?



Answers *



Dairy products



Fruits and vegetables



Meat poultry, fish, eggs



Cereals and baked goods

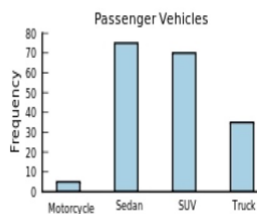
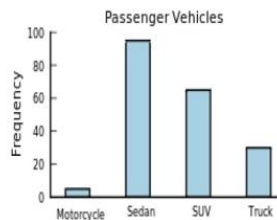
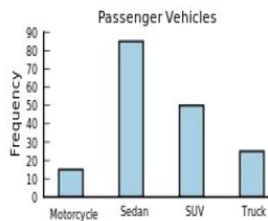
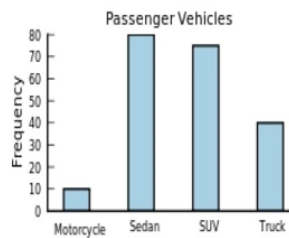
Problem 20

The following frequency distribution presents the frequency of passenger vehicles that pass through a certain intersection from 8:00 AM to 9:00 AM on a particular day.

Vehicle Type	Frequency
Motorcycle	5
Sedan	75
SUV	70
Truck	35

Construct a frequency bar graph for the data.

Answers *



Problem 21

Choose the answer below that best completes the following statement.

A _____ is a number that describes a sample.

Answers *

☐

measurement

☐

population

☒

statistic

☐

parameter

The sum of the relative frequencies for all classes will always equal

Answers *

☒

one

☐

the number of classes

☐

the number of items in the study

☐

100

Problem 24

Let Z be the standard normal random variable. Given that $P(Z < Z_0) = 0.758$, what is Z_0 ?

Answers *

☐

0.750

☒

0.700

☐

0.242

☐

-0.70

1. What to Find?

- ☐ Probability (P_0)
☒ Percentile (X_0)

2. X_0 in Which Probability?

- ☐ $P[X_0 < X < V] = P_0$
☐ $P[V < X < X_0] = P_0$
☐ $P[X > X_0] = P_0$
☒ $P[X < X_0] = P_0$

Given Probability: P_0

0.758

3. Input Information

Population Mean: μ

0

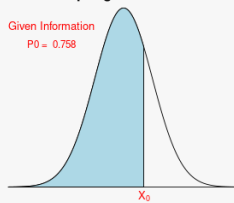
Population Standard Deviation: σ

1



Sampling Distribution

Given Information
 $P_0 = 0.758$



Question: Given $P(X < X_0) = 0.758$, what is X_0 ?

Step 3. Finding X_0 :

$$P\left(\frac{X - 0}{1} < \frac{X_0 - 0}{1}\right) = 0.758$$

or equivalently,

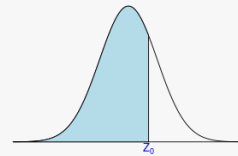
$$P\left(Z < \frac{X_0 - 0}{1}\right) = 0.758$$

Therefore,

$$\frac{X_0 - 0}{1} = Z_0 = 0.7.$$

Step 4. Answer: $X_0 = 0 + (0.7) \times 1 = 0.7$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 0}{1}.$$

Step 2. Find the Z-score corresponding to X_0

$$P(Z < Z_0) = 0.758.$$

Therefore, $Z_0 = 0.7$.

Problem 25

Tomkins Associates reports that the mean clear height for a Class A warehouse in the United States is 22 feet. Suppose clear heights are normally distributed and that the standard deviation is 4 feet.

Find the clear height such that 10% of all clear heights are less than it.

Answers *

☐

0.9

☐

1.28

☐

-1.28

☒

16.88

1. What to Find?

- ☐ Probability (P_0)
☒ Percentile (X_0)

2. X_0 in Which Probability?

- ☐ $P[X_0 < X < V] = P_0$
☐ $P[V < X < X_0] = P_0$
☐ $P[X > X_0] = P_0$
☒ $P[X < X_0] = P_0$

Given Probability: P_0

0.1

3. Input Information

Population Mean: μ

22

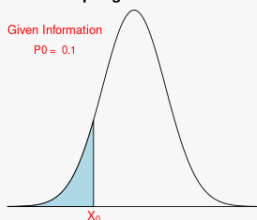
Population Standard Deviation: σ

4



Sampling Distribution

Given Information
 $P_0 = 0.1$



Question: Given $P(X < X_0) = 0.1$, what is X_0 ?

Step 3. Finding X_0 :

$$P\left(\frac{X - 22}{4} < \frac{X_0 - 22}{4}\right) = 0.1$$

or equivalently,

$$P\left(Z < \frac{X_0 - 22}{4}\right) = 0.1$$

Therefore,

$$\frac{X_0 - 22}{4} = Z_0 = -1.28.$$

Step 4. Answer:

$$X_0 = 22 + (-1.28) \times 4 = 16.88$$

Standard Normal Distribution



Step 1. Z-score Transformation

$$Z = \frac{X - 22}{4}.$$

Step 2. Find the Z-score corresponding to X_0

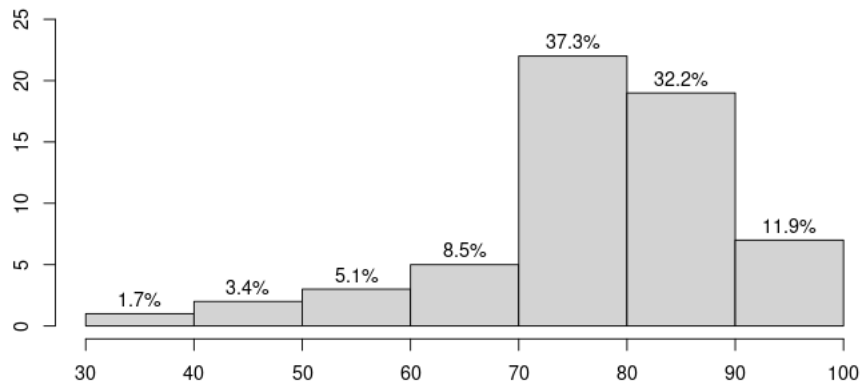
$$P(Z < Z_0) = 0.1.$$

Therefore, $Z_0 = -1.28$.

Summary of Midterm #1

The overall performance of the class in midterm #1 is good (see the following summary). Please keep working hard for the next exam. The materials to be covered in midterm #2 will be slightly more technical. Please study the weekly materials seriously and use the apps to verify your manual work and make sure you understand the concepts and procedures clearly!

Probability Distribution Histogram



1. Five Number Summary :

The five-number summary is used to describe the shape of the distribution of a given numerical data. It consists of five numbers: minimum data value, first quartile, median, the third quartile, and the maximum data value.

The five-number summary of this given data set is:

stats	value
Min.	36.00
1st Qu.	72.00
Median	80.00
3rd Qu.	88.00
Max.	100.00

2. Boxplot :

The boxplot is a geometric representation of the five-number summary. The boxplot of the given data set is given below.

