Research Experience for Teachers: Mitigating Natural Disasters

NDSU | NORTH DAKOTA STATE UNIVERSITY

Lesson #4 The Normal Distribution

Subject Area	Mathematics
Grade Level	9-12
Prior knowledge	Measures of center and spread for univariate data
Time required	1 class period

Summary

Many random variables in the real world follow the normal distribution. The normal distribution is completely determined by its mean and its standard deviation. The distribution of certain sample statistics such as the sample mean are approximately normally distributed when the sample size is relatively large. For this reason, the normal distribution is primary to our ability to use sampling to make inferences about a population. In this lesson students will use the Empirical Rule to label the axis of a normal distribution as well as calculate percentages and probability.

Education Standard

NCTM Principals and Standards

Select and use appropriate statistical methods to analyze data

- Find, use, and interpret measures of center and spread, including mean and interquartile range.
- For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.

I can statement

I can analyze data that is normally distributed.

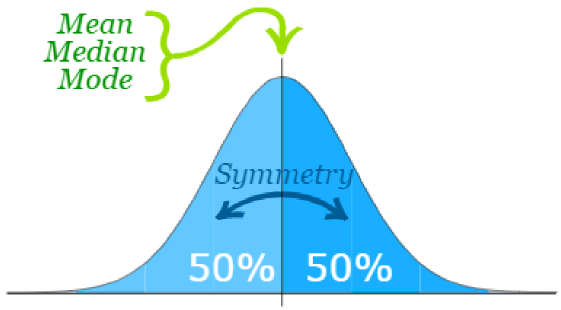
Introduction

Dr. Nic's Math and Stats video: *Understanding the Normal Distribution*

https://creativemaths.net/videos/video-distributions-normal/

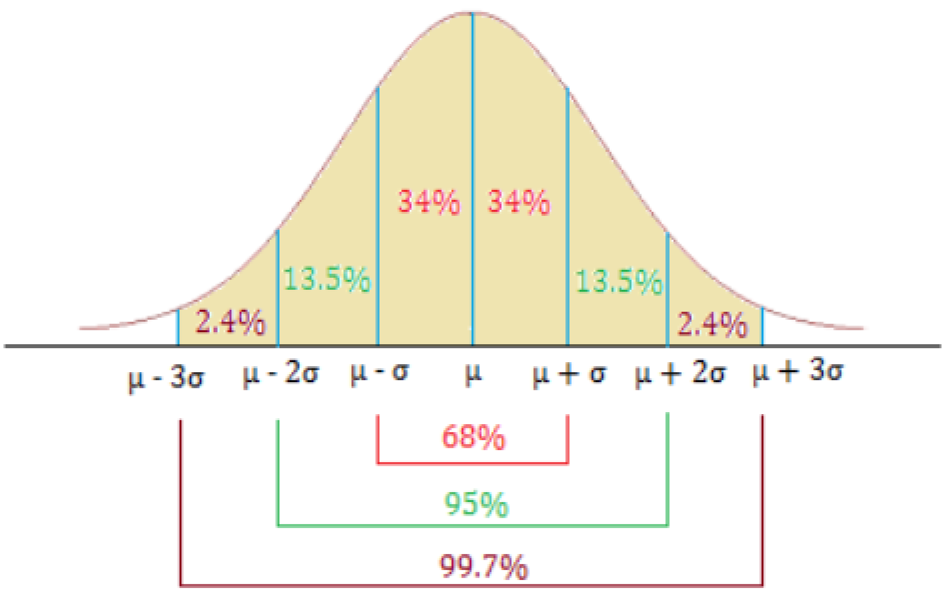
Notes and Examples

A **Normal Distribution** has a symmetric bell shape centered at the **Mean**.



The **Empirical Rule** describes the spread of a Normal Distribution.
If data is normally distributed, then

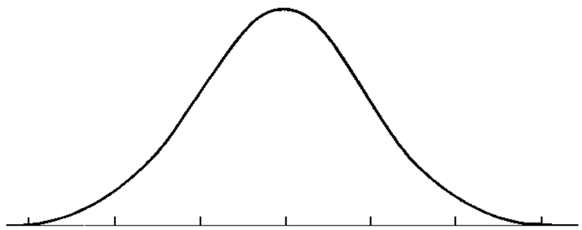
> **68%** of the observations are within **one** standard deviation of the mean.
> **95%** of the observations are within **two** standard deviation of the mean.
> **99.7%** of the observations are within **three** standard deviation of the mean.



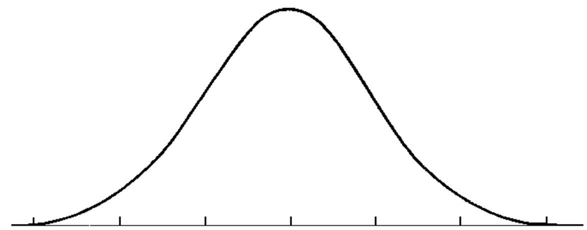
Empirical rule

1. Label the x-axis of each **Normal Distribution** at one, two, and three **Standard Deviations** from the **Mean**.

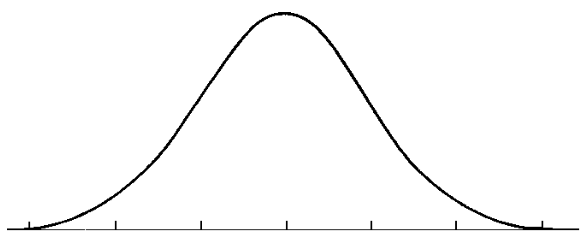
a) mean = 45, standard deviation = 5



b) mean = 95, standard deviation = 2



2. The scores on an exam are normally distributed, with a mean of 85 and standard deviation of 5.



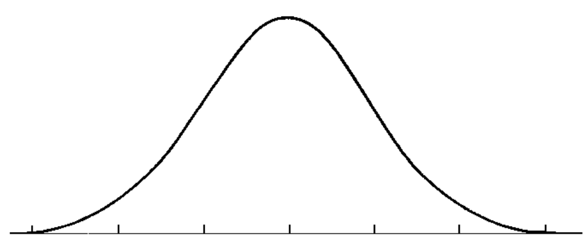
What percent of the scores are between 85 and 95?

____

What is the probability that a randomly selected exam has a score less than 75%?

____

3. The heights of adult American males are normally distributed with a mean height of 69.5 inches and standard deviation of 2.5 inches.



___________ % of adult American males are between 67 and 72 inches tall.

___________ % of adult American males are between 64.5 and 74.5 inches tall.

___________ % of adult American males are between 62 and 77 inches tall.

___________ % of adult American males are between 67 and 74.5 inches tall.

How many men out of a random sample of size n=1000 would you expect to be taller than 72 inches

Author
Martha J. Nelson

Acknowledgements
This curriculum was developed under the National Science Foundation RET grant #1953102. However, these contents do not necessarily represent the policies of the National Science Foundation, and you should not assume endorsement by the federal government.

Exercise #4 The Normal Distribution

The Data:

The data in for this exercise was collected from a model of the intersection of Labeaux and County Road 18 in Albertville MN created using PTV VISSIM software. Videos of traffic simulation which generated the data:

0% AV saturation: https://www.youtube.com/watch?v=IkwWuVd-EAI
90% AV saturation: https://www.youtube.com/watch?v=S3Czr_taME0
3D view: https://www.youtube.com/watch?v=ylzQreGefqY

The Variables:

The software collected data on the following variables:

Average Vehicle Delay is the average number of seconds a vehicle is stopped at the intersection.

Average Queue Length is the average length in meters of the line of vehicles stopped.

Maximum Queue Length is the maximum length in meters of the line of vehicles stopped.

Number of Queue Stops is the number of stops made by all vehicles at the intersection.

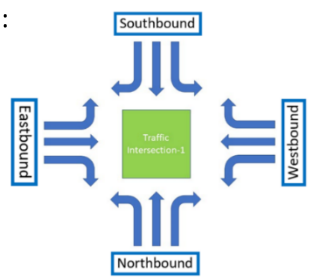
These variables are measured at each entrance to the intersection:

Westbound **(WB)**

Southbound **(SB)**

Eastbound **(EB)**

Northbound **(NB).**



Each simulation lasted 90 minutes (5400 seconds) with data collected every 15 minutes (900 seconds) and the first and last 15-minute intervals were discarded. Therefore, the data is collected for four time intervals for each run of the simulation:

900-1800 seconds, **1800-2700** seconds, **2700-3600** seconds, **3600-4500** seconds

Each variable is measured for seven autonomous vehicle saturation rates:

0% AV – all human driven cars

15% AV – 85% human driven cars

30% AV – 70% human driven cars

45% AV – 55% human driven cars

60% AV – 40% human driven cars

75% AV – 25% human driven cars

90% AV – 10% human driven cars

At each AV saturation rate, the simulation was run ten times. As a result, at each variable at each AV saturation rate there are 160 observations. One for each of the four directions at each of the four time intervals for each of the ten simulations.

At each saturation rate, the autonomous vehicles are tested using three different driving behaviors:

Cautious, **Normal**, **Aggressive**

Queue stops

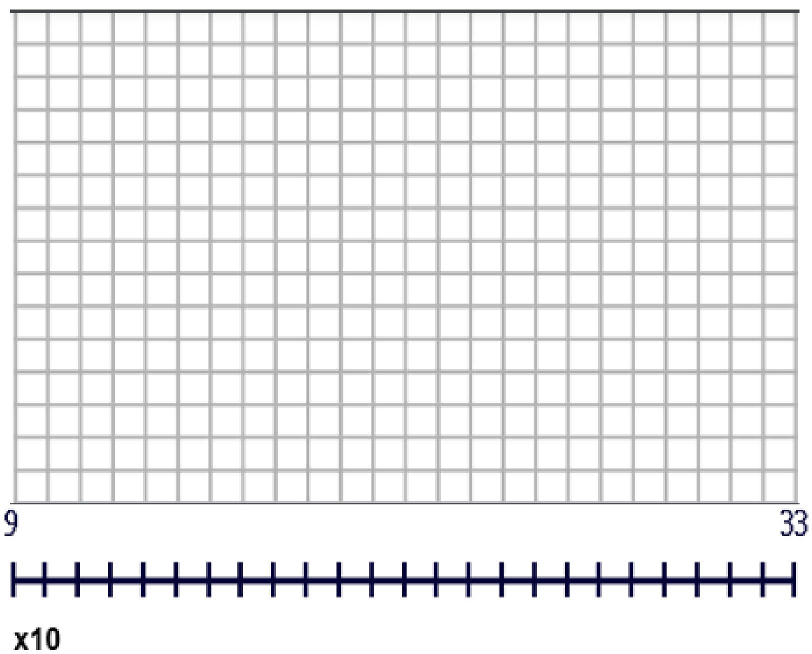
0% AV

Run	TIMEINT	WB	QSTOPS	SB	QSTOPS	EB	QSTOPS	NB	QSTOPS
1	900-1800	1	123	2	312	3	113	4	274
1	1800-2700	1	104	2	299	3	100	4	279
1	2700-3600	1	127	2	234	3	124	4	297
1	3600-4500	1	157	2	247	3	142	4	228
2	900-1800	1	139	2	258	3	173	4	274
2	1800-2700	1	184	2	258	3	116	4	269
2	2700-3600	1	152	2	285	3	119	4	263
2	3600-4500	1	141	2	303	3	102	4	208
3	900-1800	1	117	2	271	3	134	4	252
3	1800-2700	1	116	2	291	3	117	4	235
3	2700-3600	1	107	2	249	3	134	4	280
3	3600-4500	1	126	2	233	3	110	4	207
4	900-1800	1	105	2	228	3	103	4	274
4	1800-2700	1	135	2	273	3	120	4	271
4	2700-3600	1	110	2	289	3	141	4	297
4	3600-4500	1	131	2	329	3	121	4	257
5	900-1800	1	176	2	267	3	122	4	301
5	1800-2700	1	138	2	248	3	131	4	262
5	2700-3600	1	167	2	280	3	128	4	278
5	3600-4500	1	141	2	312	3	139	4	289
6	900-1800	1	128	2	254	3	117	4	237
6	1800-2700	1	104	2	243	3	102	4	294
6	2700-3600	1	145	2	301	3	95	4	249
6	3600-4500	1	157	2	246	3	124	4	294
7	900-1800	1	107	2	311	3	139	4	280
7	1800-2700	1	148	2	262	3	135	4	320
7	2700-3600	1	113	2	289	3	102	4	253
7	3600-4500	1	132	2	275	3	108	4	267
8	900-1800	1	148	2	238	3	120	4	285
8	1800-2700	1	130	2	252	3	165	4	294
8	2700-3600	1	143	2	306	3	129	4	250
8	3600-4500	1	122	2	285	3	158	4	255
9	900-1800	1	141	2	240	3	150	4	246
9	1800-2700	1	133	2	264	3	144	4	275
9	2700-3600	1	140	2	269	3	145	4	328
9	3600-4500	1	143	2	233	3	120	4	256
10	900-1800	1	125	2	244	3	165	4	257
10	1800-2700	1	135	2	278	3	144	4	258
10	2700-3600	1	132	2	265	3	122	4	289
10	3600-4500	1	134	2	277	3	127	4	239

Using the number of queue stops data at 0% AV for each intersection entrance direction (WB, SB, EB, NB):

1. Sort the data in ascending order.
2. Construct a histogram. (You may copy or use the histogram you constructed in lesson #1)
3. Report the mean **($\bar{x}$)** and standard deviation **(s)**.
4. Calculate the intervals that capture 1, 2, and 3 standard deviations above and below the mean.
5. Calculate the percentage of observations that fall within each of the three intervals.

Westbound



$\bar{x}$ = ____________ s = ____________

($\bar{x}$ - s , $\bar{x}$ + s): (__________, __________)

________% of observations

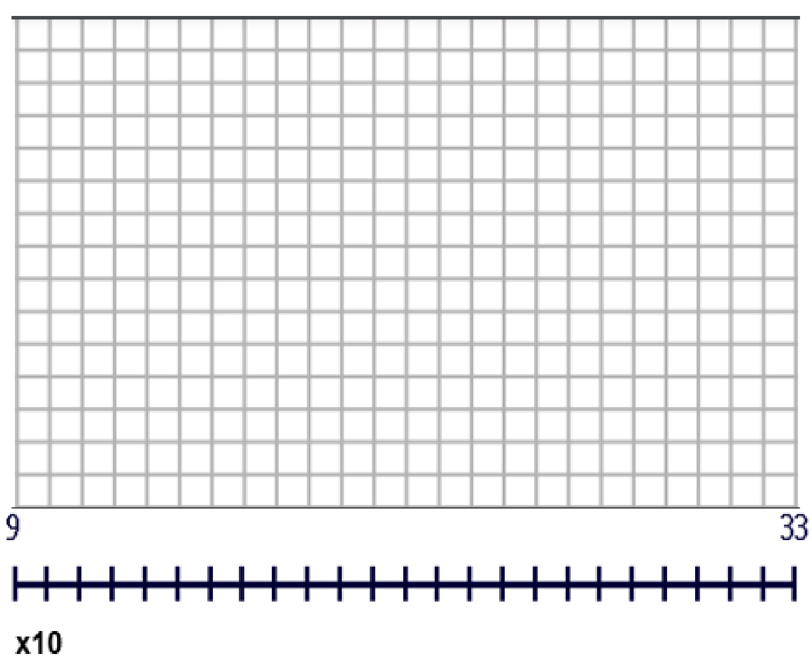
($\bar{x}$ - 2s, $\bar{x}$ +2s): (__________, __________)

________% of observations

($\bar{x}$ -3s, $\bar{x}$ + 3s): (__________, __________)

________% of observations

Southbound



$\bar{x}$ = ___________ s = ____________

($\bar{x}$ - s , $\bar{x}$ + s): (__________, __________)

________% of observations

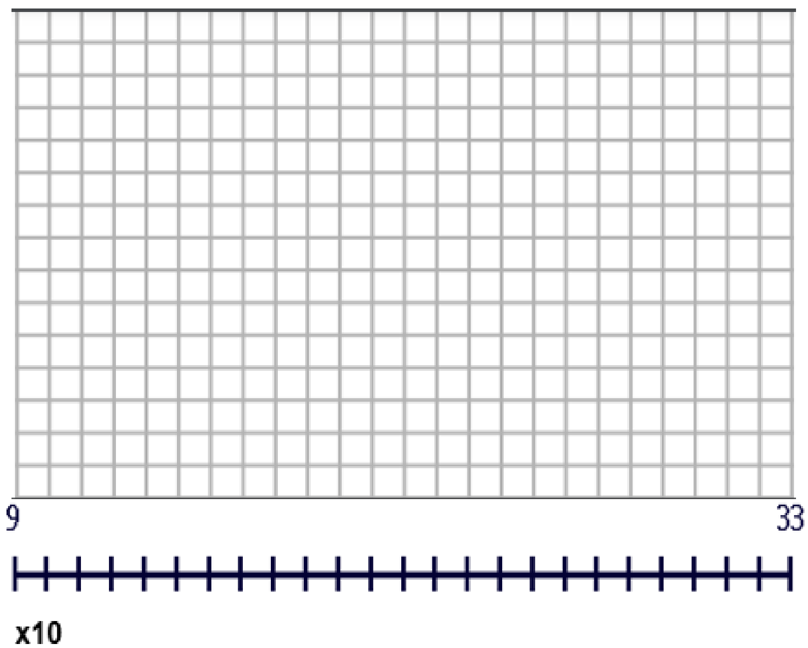
($\bar{x}$ - 2s , $\bar{x}$ +2s): (__________, __________)

________% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (__________, __________)

________% of observations

Eastbound



$\bar{x}$ = __________ s = __________

($\bar{x}$ - s , $\bar{x}$ + s): (_________, _________)

________% of observations

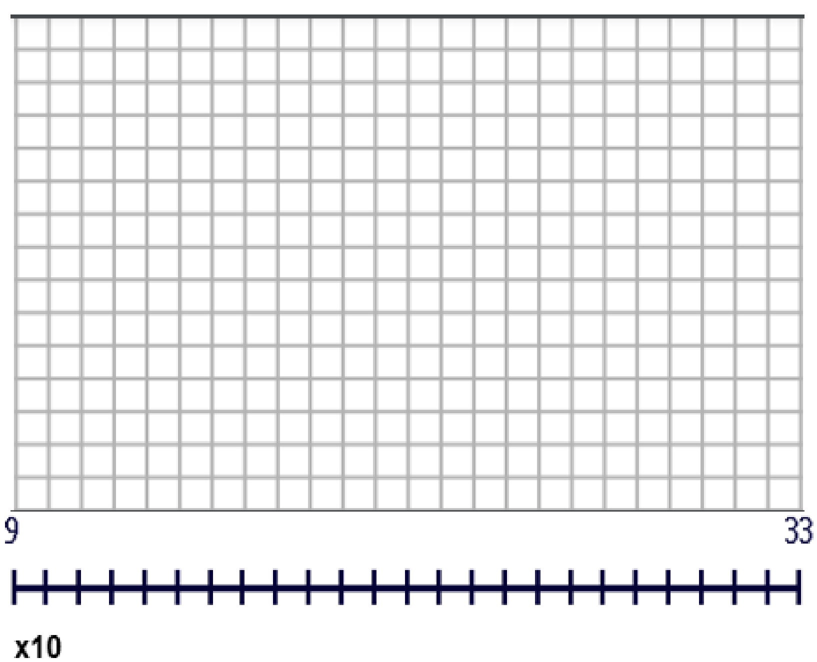
($\bar{x}$ - 2s , $\bar{x}$ +2s): (_________, _________)

________% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (_________, _________)

________% of observations

Northbound



$\bar{x}$ = __________ s = __________

($\bar{x}$ - s , $\bar{x}$ + s): (_________, _________)

________% of observations

($\bar{x}$ - 2s, $\bar{x}$ +2s): (_________, _________)

________% of observations

($\bar{x}$ -3s, $\bar{x}$ + 3s): (_________, _________)

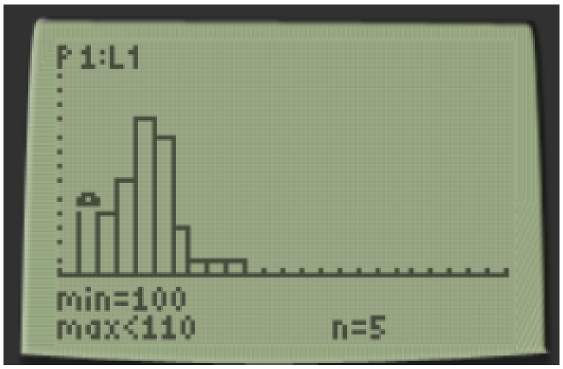
________% of observations

Exercise#4 The Normal Distribution Answers

Using the number of queue stops data at 0% AV for each intersection entrance direction:

1. Sort the data in ascending order.
2. Construct a histogram. (You may copy or use the histogram you constructed in lesson #2)
3. Report the mean **($\bar{x}$)** and standard deviation **(s)**.
4. Calculate the intervals that capture 1, 2, and 3 standard deviations above and below the mean.
5. Calculate the percentage of observations that fall within each of the three intervals.

Westbound



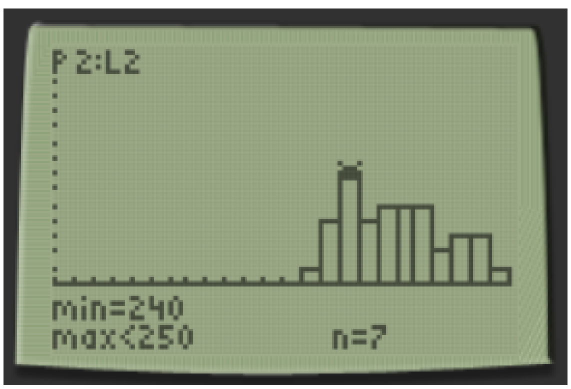
$\bar{x}$ = 133.9 s = 18

($\bar{x}$ - s , $\bar{x}$ + s): (115.03 , 152.78) 70% of observations

($\bar{x}$ - 2s , $\bar{x}$ +2s): (96.15 , 171.65) 95% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (77.3 , 190. 53) 100% of observations

Southbound



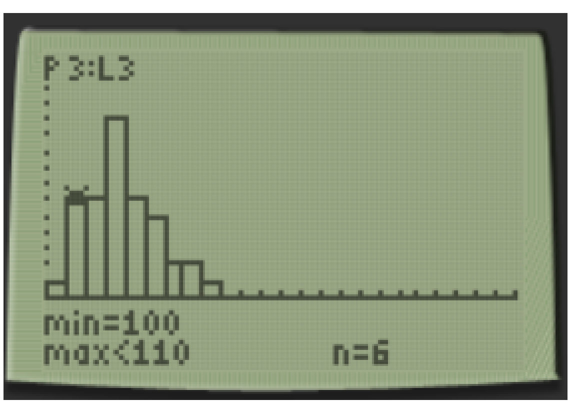
$\bar{x}$ = 269.95 s = 26.26

($\bar{x}$ - s , $\bar{x}$ + s): (243.7 , 296.21) 62.5% of observations

($\bar{x}$ - 2s , $\bar{x}$ +2s): (217.4 , 322.5) 97.5% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (191.2 , 348.7) 100% of observations

Eastbound



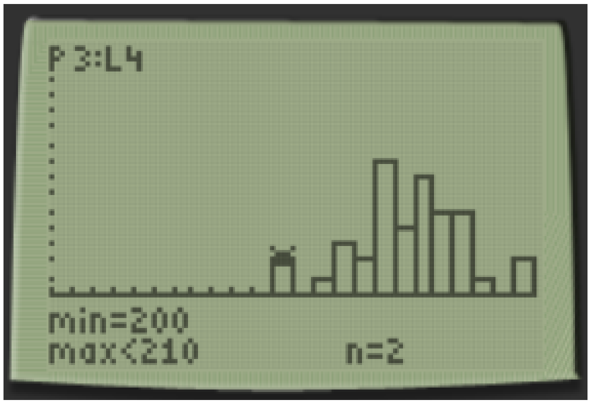
$\bar{x}$ = 127.5 s = 18.9

($\bar{x}$ - s , $\bar{x}$ + s): (108.6 , 146.4) 70% of observations

($\bar{x}$ - 2s , $\bar{x}$ +2s): (89.7 , 165.3) 95% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (70.8 , 184.2) 100% of observations

Northbound



$\bar{x}$ = 268.025 s = 26.4

($\bar{x}$ - s , $\bar{x}$ + s): (241.6 , 294.4) 72.5% of observations

($\bar{x}$ - 2s , $\bar{x}$ +2s): (215.2 , 320.8) 92.5% of observations

($\bar{x}$ -3s , $\bar{x}$ + 3s): (188.8 , 347.2) 100% of observations