Chapter 3 Averages & Variation

Measures of center and variation are helpful summaries of how a larger data set behaves.

Notation

IMPORTANT
Always refer to population or sample when describing mean, var, or std. dev.

Measure	Population Parameter	Sample Statistic
mean	μ ← (greek "mu")	$\bar{x}$
variance	σ^2 ← (greek "sigma^2")	s^2
standard deviation	σ ← (greek "sigma")	s

Note
If you know the variance you can calculate the standard deviation and vice versa.

variance = (std. dev.)^2 and std. dev. = $\sqrt{\text{variance}}$

Calculator

"1 VAR STATS" is a calculator function that takes in a list of data and (if applicable) a frequency list and outputs the following:

$\bar{x}$: population or sample mean (BAD NOTATION! But the calculator doesn't know if you entered a pop. or sample)
Σx : sum of all data values
Σx^2 : sum of all (data values)^2
Sx : sample std. dev. = s
σx : pop. std. dev. = σ
n : size of dataset
minX : minimum data value
Q_1 : 1st quartile
med : median
Q_3 : 3rd quartile
maxX : maximum data value

HOW TO: TI-nspire see Resources

① Enter data: STAT ⇒ EDIT
Enter x data into L_1 and if there is a frequency list into L_2.

② Calculate: STAT ⇒ CALC ⇒ 1Var Stats
In "List" enter L_1 or wherever you entered data.
Leave "FreqList" blank unless you entered frequency in a list.

Data Sets

Set A: {10, 9, 9, 8, 8, 8, 8, 7, 7, 6, 6, 6, 6, 5, 5, 4}
Set B: {10, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 4}

Alternate Way to Represent Data:

Set A

X	Freq.
4	1
5	2
6	4
7	2
8	4
9	2
10	1

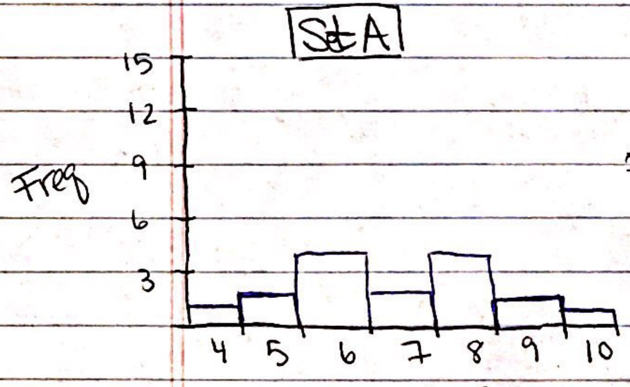
Set B

X	Freq.
4	1
7	14
10	1

Histogram Shape:

- Mean = median
  Symmetric
- Mean < Median
  Left Skewed
- Mean > Median
  Right Skewed

Histograms:



Measures of Center:

Set A

- No mode (if there is a tie, then neither is the mode).
- Median = 7
- Mean = 7

Set B

- Mode = 7
- Median = 7
- Mean = 7

Both Symmetric

NOTE: All measures of center measure a different meaning of "center".

Measures of Variation:

Set A

population { var. = 2.624
std. dev. = 1.620

sample { var. = 2.799
std. dev. = 1.673

Set B

population { var. = 1.125
std. dev. = 1.061

sample { var. = 1.199
std. dev. = 1.095

NOTE: We did not specify if these sets were a population or sample so both are calculated.

Calculator Try to confirm the values of the measures of center and variation to check if you understand how to use the calculator function "1 VarStats".

Comment: Without seeing the data and the histograms it would have been difficult to know that Set A | Set B are drastically different.

5 Number Summary

Partitions the data into quartiles.

Data: 25% | 25% | 25% | 25%

min — Q_1 — median — Q_3 — max

Set A: $\underbrace{4\ 5\ 5\ 6}_{25\%}\ \underbrace{6\ 6\ 6\ 7}_{25\%}\ \underbrace{7\ 8\ 8\ 8}_{25\%}\ \underbrace{8\ 9\ 9\ 10}_{25\%}$

5 Number Summary

min = 4

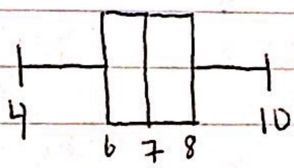
$Q_1 = 6$

med = 7

$Q_3 = 8$

max = 10

Box & Whisker Plot



Chebyshev's Theorem

- Gives a theoretical lower bound for how spread the data is.
- Applies to any data set.

To make a conclusion via Chebyshev's Theorem:

① Know the mean & std. dev. of the data set.
* Applies to both population or sample.

② Fix how many std. devs from center (mean) you want a lower bound for. We denote this by "k". k must be strictly more than 1.

③ Conclude:
Atleast $(1-1/k^2)\times 100\%$ of the data points are within k std. dev.s of the mean.

Example) Assume Set A is a sample.

Set A
$\bar{x} = 7$
$s = 1.673$

Q: According to Chebyshev's theorem, how much of the data is within $k=1.5$ std. dev.s of the mean?

A: $(1-1/(1.5)^2)\times 100 = 55.6\%$
Atleast 55.6%.

Q: What is the actual % of data with 1.5 std. devs of the mean for Set A?

A: 1.5 std. devs below mean:
$\bar{x} - 1.5(s) = 7 - 1.5(1.673) = 4.49$
1.5 std. devs above mean:
$\bar{x} + 1.5(s) = 7 + 1.5(1.673) = 9.51$

Find % of data between 4.49 and 9.51:
14 out of 16 $\Rightarrow$ $(14/16)\times 100 = 87.5\%$

Take Away: We knew atleast 55.6% of data was between 4.49–9.51, in reality there is $87.5\% \geq 55.6\%$.