Part I (Section 2) Statistics

Descriptive Statistics

B) Measure of Positions (**Q**uartiles, **B**ox-Plot, **P**ercentile, **Z**-score)

Quartiles: Breaking the **ranked data** in 3 quartiles (**Q1, Q2, Q3**)

Data: ______25%_______**Q1**_______ 25%_________**Q2**________25%_______**Q3**_______25%________

How to find quartiles? 3 steps
Rank the data points, Find **Q2** = Median and the new medians **Q1, Q3** on either side of Q2.

Example 1: **Odd** number of data Data: 2, 5, 11, 16, 8, 9, 3, 7, 5, 4, 13

Ranked Data: 2, 3, ,**4**, 5, 5, **7**, 8, 9, **11**, 13, 16,
Q1 Q2 Q3

Example 2: **Even** number of data points Data: 2, 3, 5, 5, 7, 8, 9, 11, 16, 4
Ranked Data 2, 3, **4**, 5, 5, 7, 8, **9**, 11, 16, **Q2 = Median** = $(5+7)/2=6$
Q1 **Q2** = 6 **Q3**

TI-83/84 *Inputting data in **L1** (stat $\rightarrow$ Option 1 $\rightarrow$ enter)*
then *stat $\rightarrow$ calc $\rightarrow$ Option 1 $\rightarrow$ enter $\rightarrow$ 2n d $\rightarrow 1$ $\rightarrow$ enter*

Extra Practice: *Answer questions on columns **A-G** on **page 3** of practice problem **part 1***

C) Measure of Variation (**R**ange, **S**tandard Deviation, **V**ariance)

Range: It shows how far apart the data points are? **Range** = the highest value - the smallest value

Standard Deviation (σ, s) : It measures the **average dispersion** of data **around the mean.**

Example: Consider the 3 random delivery time (in days) taken by 2 different companies A, and B

	A	**B**
Mean	5	5
Median	5	5
Mode	5	none

At first it seems there are not that much of difference between the delivery times of these two companies but let's look at their actual data and their plots on Dot-Plot.

	A	B
Delivery time	5	5
Delivery time	5	0
Delivery time	5	10

A **Dot Plot** **B**

```
A:            x
              x
____  ____x________
0         5        10

B:
x________x________x
0        5        10
```

Now, it seems that there is **no dispersion** for company A, but an **average dispersion of 5** for company B, suggesting that company is more reliable meeting the average delivery time.

The formula for the Standard Deviation or average dispersion of data around mean $s = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n-1}}$

Company A

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
5	5	0	0
5	5	0	0
5	5	0	0
			$\sum (x-\bar{x})^2 = 0$

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{0}{3-1}} = \sqrt{0} = 0$$

Company B

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
5	5	0	0
0	5	-5	25
10	5	5	25
			$\sum (x-\bar{x})^2 = 50$

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{50}{3-1}} = \sqrt{25} = 5$$

Find the mean and standard deviation for 5, 6, 3 , 9, 10, 3, and also draw the **dot-plot**.

x	$\bar{x} = \frac{\sum x}{n} = \frac{36}{6} = 6$	$(x-\bar{x})$	$(x-\bar{x})^2$
5	6	-1	1
6	6	0	0
3	6	-3	9
9	6	3	9
10	6	4	16
3	6	-3	9
$\sum x =$			$\sum (x-\bar{x})^2$ =44

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{44}{6-1}} = \sqrt{8.8} = 2.96 \ \ 2.97$$

Variance = $s^2 = 8.8$

Variance (σ^2 , s^2): Variance is the **square of standard deviation.**

TI-83/84 *Inputting data in* ***L1*** *(stat $\rightarrow$ Option 1 $\rightarrow$ enter)*

then *stat $\rightarrow$ calc $\rightarrow$ Option 1 $\rightarrow$ enter $\rightarrow$ 2nd $\rightarrow$ 1 $\rightarrow$ enter*

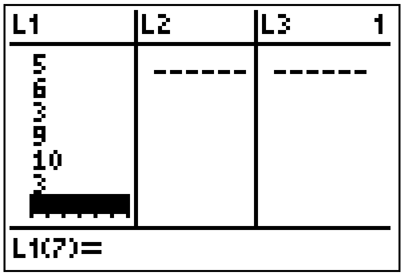
Rule of thumb to **estimate s**: $s = \dfrac{Range}{4}$ Generally the larger the data set the closer the estimate will be to the exact value.

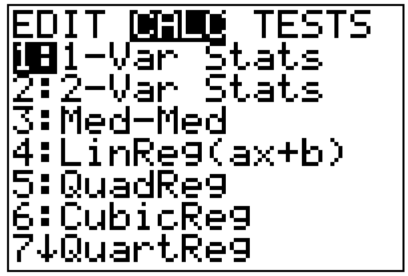
Extra Practice: *Answer questions on columns* ***A-G*** *on* ***page 3*** *of practice problem* ***part 1***

TI-83/84

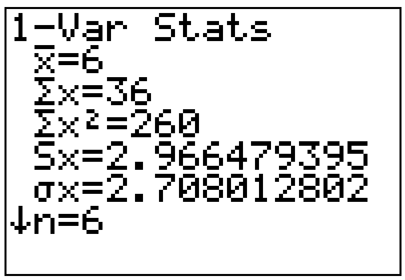
Find the mean, median, Q1, Q3 and standard deviation for 5, 6, 3 , 9, 10, 3, and also draw the Box-Plot.

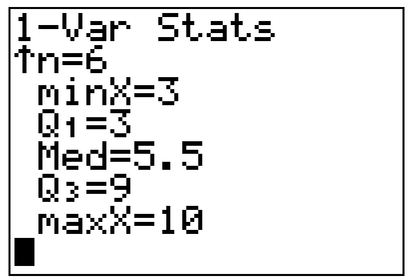
*Inputting data in **L1** (stat → Option 1 → enter)*   *stat → calc → Option 1 → enter*





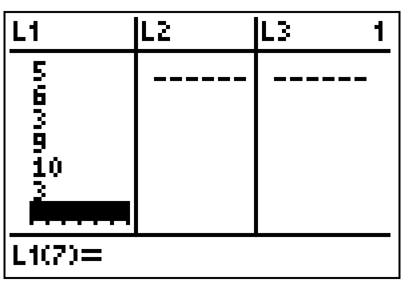
2n d → 1   **enter**   ***Results***   ***Use down arrow for more Results***

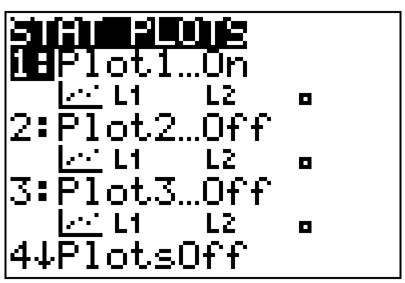


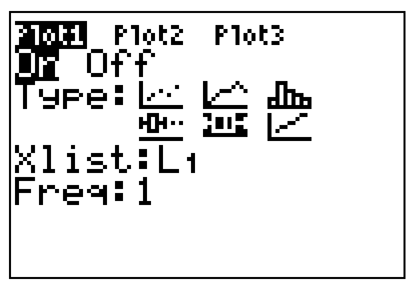


Doing the Box Plot by TI

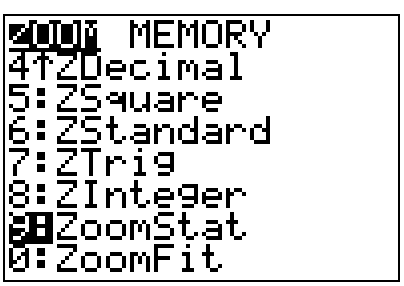
Inputting data in L1   ***2nd STAT Plots***   ***Choose the fifth option***

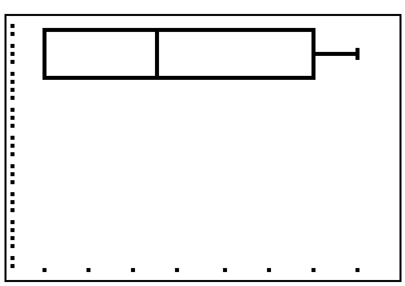






Press ZOOM 9   ***Result***





Empirical Rules: If and only if the **box-plot or histogram is centered** then we can apply the **three** following empirical rules.

$99.7\% = \bar{x} \pm 3\,s$ **99.7 %** of data are within $3\,s$ of the mean ($\bar{x}$)

$95\% = \bar{x} \pm 2\,s$ **95 %** of data are within $2\,s$ of the mean ($\bar{x}$)

$68\% = \bar{x} \pm \ s$ **68 %** of data are within $1\,s$ of the mean ($\bar{x}$)

Example: Find all three empirical rules for Abe Stat class if the average was 72 and the standard deviation was 8, assuming that Box-plot was centered.

$99.7\% = 72 \pm 3(8) = 72 \pm 24$ $48 <$ **99.7 %** of class got scores < 96

$95\% = 72 \pm 2(8) = 72 \pm 16$ $56 <$ **95 %** of class got scores < 88

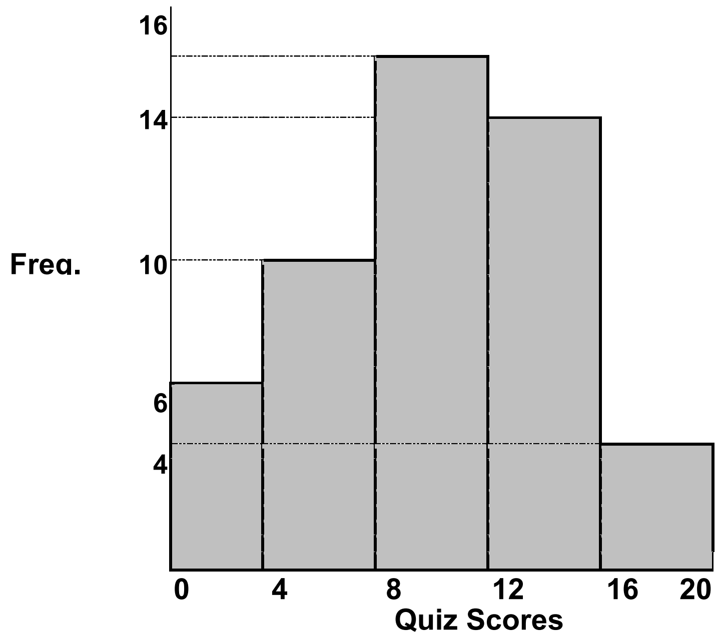
$68\% = 72 \pm 1(8) = 72 \pm 8$ $64 <$ **68 %** of class got scores < 80

Grouped Data (Freq. Table)

The table below shows the quiz scores of 50 students that are given in group.

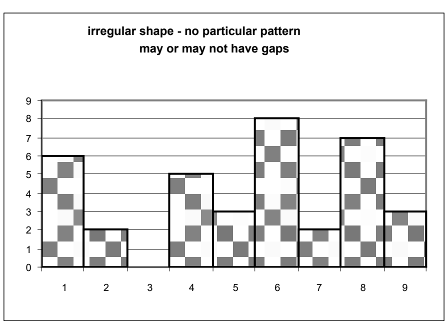
Quiz Score	Freq (f) = Students			
0 – 4	6			
4 - 8	10			
8 – 12	16			
12 – 16	14			
16 – 20	4			

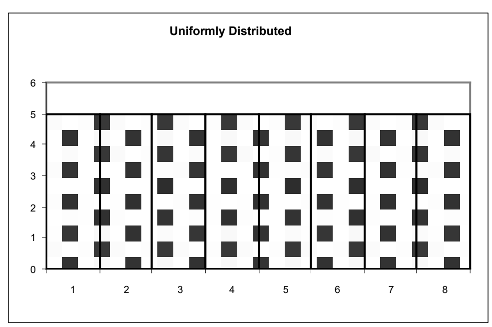
Use the quiz scores on x-axis, frequency on the Y-axis to draw blocks for a shape that is called **Histogram**

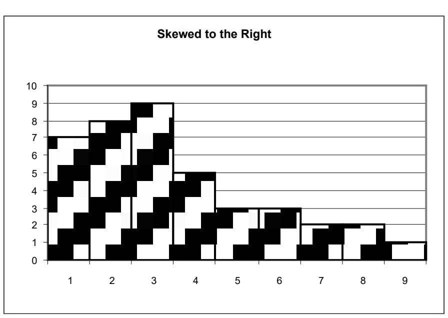


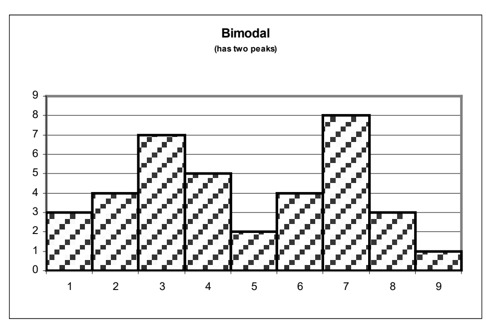
Histogram looks close to a Centered or bell-shaped distribution.

Different possible shapes of Histogram









Mean and **Standard Deviation**.

First step is to create a new column called **midpoint** (average of scores in each group). For example for 0 – 4, the midpoint will be 2, for 4 – 8, the midpoint will be 6. Next step is to open two new columns $f \times m$ and $f \times m^2$ do the necessary calculations, find the summation for each and then use them in the given formulas.

X-axis		**midpoint**	**Mean**	**St.Dev.**
Quiz Scores	**Freq(f)= Students**	m	$f \times m$	$f \times m^2$
0 – 4 4 - 8 8 – 12 12 – 16 16 – 20	6 10 16 14 4	$(0+4)/2=2$ $(4+8)/2=6$ $(8+12)/2=10$ $(12+16)/2=14$ $(16+20)/2=18$	$6\times 2=12$ $10\times 6=60$ $16\times 10=160$ $14\times 14=196$ $4\times 18=72$	$6\times 2^2=24$ $10\times 6^2=360$ $16\times 10^2=1600$ $14\times 14^2=2744$ $4\times 18^2=1296$
	$\sum f = n = 50$		$\sum f\times m=500$	$\sum f\times m^2=6024$

Mean: $\bar{X}=\dfrac{\sum f\times m}{n}=\dfrac{500}{50}=10$

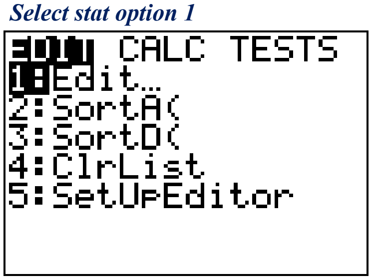
Standard deviation: $s=\sqrt{\dfrac{n\sum f\times m^2-\left(\sum f\times m\right)^2}{n(n-1)}}=\sqrt{=\dfrac{50(6024)-(500)^2}{50(50-1)}}=\sqrt{\dfrac{51200}{2450}}=4.57$

Variance: $s^2=4.57^2=20.9$

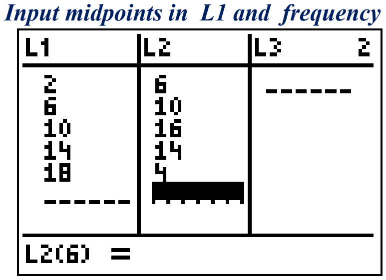
Apply 95% empirical rule: $95\% = \bar{x}\pm 2\,s = 10\pm 2(4.57)=10\pm 9.14$ $\quad 0.86 <$ **95. %** of class got scores <19.14

TI-83/84

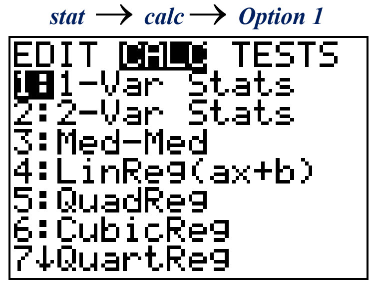
Select stat option 1



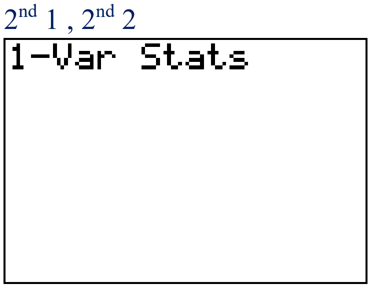
Input midpoints in L1 and frequency in L2



stat → calc → Option 1

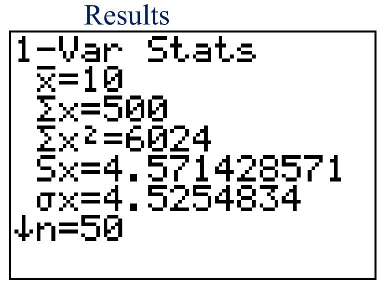


2nd 1 , 2nd 2



Press enter

Results



Practice 1: Use both formula and the Ti to find the mean, standard deviation and the variance.

Quiz Scores	**Freq(f)**	m	$f \times m$	$f \times m^2$
0 – 10 10 - 20 20 – 30 30 – 40	8 12 14 6	5 25 	40 180 	200 7350
	$\sum f = n = 40$		$\sum f \times m =$ **780**	$\sum f \times m^2 =$ **19000**

Mean: $\bar{X} = \dfrac{\sum f \times m}{n} = \underline{\qquad\qquad} = 19.5$

Standard deviation: $s = \sqrt{\dfrac{n\sum f \times m^2 - \left(\sum f \times m\right)^2}{n(n-1)}} = \sqrt{\dfrac{\qquad}{\qquad}} = \sqrt{\dfrac{\qquad}{\qquad}} = 9.86$

Variance: $s^2 = 9.8^2 = 97.18$

Apply 95% empirical rule:

Practice 2: Use both formula and the Ti to find the mean, standard deviation and the variance

Test Scores	**Freq (f)=**	m	$f \times m$	$f \times m^2$
0 – 20	2	10	20	200
20 – 40	8	30	$8 \times 30 = 240$	$8 \times 30^2 = 7200$
40 – 60	14			
60 – 80	32			
80 – 100	24			
	$\sum f = n =$		$\sum f \times m =$	$\sum f \times m^2 =$

Mean: $\bar{X} = \dfrac{\sum f \times m}{n} = \underline{\qquad\qquad} = 67$

Standard deviation: $s = \sqrt{\dfrac{n\sum f \times m^2 - \left(\sum f \times m\right)^2}{n(n-1)}} = \sqrt{\dfrac{\qquad}{\qquad}} = \sqrt{\dfrac{\qquad}{\qquad}} = 20.89$

Variance: $s^2 =$

Apply 68% empirical rule: