One-Variable Statistics

by Nathan Oldridge
(ChemistNate)

Mean / Median / Mode / Weighted Mean

1. Calculate the mean, median and mode for each set of data.

a) 2, 2, 4, 6, 10, 36

b) 2, 5, 7, 11, 11, 11, 12, 12, 13, 13, 13

c)

Number of Tickets Bought	Frequency
2	10
3	6
4	8
5	3
6	5
8	4

d)

Height (cm)	Number of Students
$120 \leq x < 130$	4
$130 \leq x < 140$	6
$140 \leq x < 150$	10
$150 \leq x < 160$	6
$160 \leq x < 170$	8

2. Four numbers have a median of 30, a mode of 50 and a mean of 29. What are the numbers?

3. The mean of six numbers, five of which are 10, 12, 18, 21 and 29, is 19. What is the sixth number?

4. Nine numbers, which are already ordered from lowest to highest, have a mean of 10, a median of 11 and a mode of 9. If the middle number of the set is increased by 1, how will each of the following change?

a) Mean:

b) Median:

c) Mode:

5. There are four versions (*"isotopes"*) of lead metal atoms. Their masses (units are *amu*) are listed below, along with how much of all lead is each type.

a) What number goes where "?" is ?

b) What is the average mass of lead?

Type of Lead	Mass (amu)	% Abundance
A	203.97	?
B	205.97	24%
C	206.98	20%
D	207.98	54%

6. To the right, are the four sections one which you were marked for your *Economics 203: Why Bitcoin is the Future* course.

Calculate your mark for the course.

Section	Contribution to Your Mark	You Scored...
Knowledge	40%	88%
Inquiry	20%	80%
Communication	10%	94%
Application	30%	84%

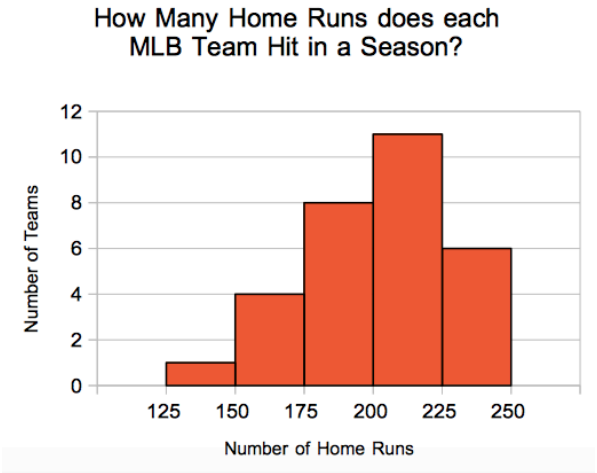
Frequency Tables and Histograms

7. The following is the height data for people in your class.
Tally the values, and find the frequency for each interval as needed for the table.
Then, create a histogram for data.
Make sure to include a title and label your axes.

121	140	134	155	127	143	138
159	162	148	153	164	131	145
162	140	149	123	142	157	129
131	149	169	160	155	145	137
160	151	143	164	137	164	

Height (cm)	Tally	Number of Students
$120 \leq x < 130$		
$130 \leq x < 140$		
$140 \leq x < 150$		
$150 \leq x < 160$		
$160 \leq x < 170$		

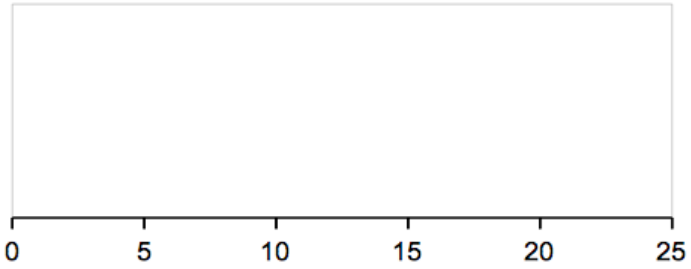
8. Convert this histogram to an interval table.



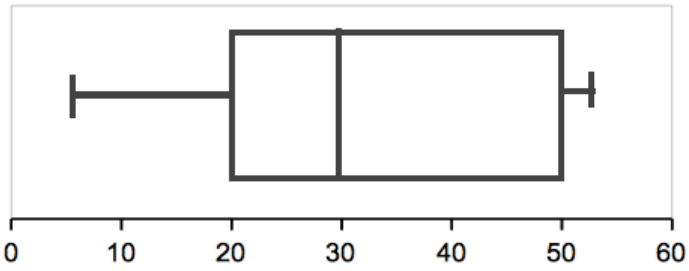
Quartiles, IQR and Box-and-Whisker Plots

9. Determine Q1, Q2 and Q3 for the following data.
Then, calculate the IQR.
Then, draw a box-and-whisker plot for it.

2
2
4
6
6
8
10
10
14
14
16
20



10. Given this box-and-whisker plot, state Q1, Q3, the IQR and the median.



Percentiles

$$p = 100\left(\frac{L + 0.5E}{n}\right) \qquad R = \frac{p}{100}(n+1)$$

11. Here are the number of questions answered correctly by 14 students on the *Chem 13 News Exam*, which has 40 questions.

16 20 22 26 26 27 29 32 34 36 36 37 38 39

a) What percentile is the student who scored 32 ?

b) What score would a student have needed to be in the 80^{th} percentile?

Variance and Standard Deviation

12. Calculate the standard deviation, and variance, of each set of data. Assume that each set of data is the *population*.

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$$

a) 2, 2, 4, 6, 10, 36

b) 2, 2, 4, 6, 8, 8

c) The data from (a) above has a range of 34, and the data from (b) has a range of 6. Does this relate to the standard deviation? Explain.

d) Recalculate the standard deviation of each data set, assuming that they are *samples* instead of populations. Note: This is **not** a long question – it is simply plugging in numbers you already know.

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

z-scores

13. The data {2, 2, 4, 6, 10, 36} has mean 10 and standard deviation of 11.94.

a) What is the z-score of each data point?

b) If a 7^{th} data point was added, with a z-score of +1.09, what would its value have been?

14. For a different set of data, which you don't have access to, a data point with x=12 has z=-1.4 and a data point with x=16 has z=-0.4. What is the mean and standard deviation of the data?

Answers

1. Calculate the mean, median and mode for each set of data.

a) 2, 2, 4, 6, 10, 36

Mean: Add them all up, and divide by how many there are: $\frac{2+2+4+6+10+36}{6}=\frac{60}{6}=10$

Median: There are two "middlest" numbers, 4 and 6: $\frac{4+6}{2}=5$

Mode: Most common data point: 2

b) 2, 5, 7, 11, 11, 11, 12, 12, 13, 13, 13

Mean: Add them all up, and divide by how many there are:
$\frac{2+5+7+11+11+11+12+12+13+13+13}{11}=\frac{110}{11}=10$. Oops! Didn't mean to make them the same. OR DID I?

Median: Cancel from either side until the middle number reveals itself: it is 11.

~~2, 5, 7,~~ ~~11, 11,~~ 11, ~~12, 12,~~ ~~13, 13, 13~~

Mode: Most common data point: 11 and 13. There can be two (or even more) modes.

c)

Number of Tickets Bought	Frequency
2	10
3	6
4	8
5	3
6	5
8	4

Mean: There are ten 2's, six 3's, eight 4's, etc...

$\frac{10(2)+6(3)+8(4)+3(5)+5(6)+4(8)}{10+6+8+3+5+4}$

$=\frac{147}{36}=4.08$

Median: There are 36 data points, so the "middlest" are numbers 18 and 19. Both are "4" so the median is 4.

Mode: 2 is the most common number (there are ten of them)

d)

Interval	Approximate Height	Number of Students
$120 \le x < 130$	125	4
$130 \le x < 140$	135	6
$140 \le x < 150$	145	10
$150 \le x < 160$	155	6
$160 \le x < 170$	165	8

Mean:

$\frac{4(125)+6(135)+10(145)+6(155)+8(165)}{4+6+10+6+8}$

$=\frac{5010}{34}=147.35$

Median: The "middlest" ones are 17 and 18. Both are 145 cm.

Mode: 145 cm (appears ten times)

2. Four numbers have a median of 30, a mode of 50 and a mean of 29. What are the numbers?

50 must appear twice. They are both of the numbers above the median.
The median of the 2nd and 3rd numbers is 30, so the 2nd number is 10.

To calculate the 1st number, we need the mean: $\frac{x+10+50+50}{4}=29$

$$x+110=116$$
$$x=116-110=6$$

3. The mean of six numbers, five of which are 10, 12, 18, 21 and 29, is 19. What is the sixth number?

$$\frac{10+12+18+21+29+x}{6}=19$$
$$90+x=114$$
$$x=114-90=24$$

4. Nine numbers, which are already ordered from lowest to highest, have a mean of 10, a median of 11 and a mode of 9. If the middle number of the set is increased by 1, how will each of the following change?

a) Mean: Goes up by a small amount, because the sum of all the numbers goes up a bit

b) Median: Goes up by 1, because the middle number went up by 1.

c) Mode: Probably doesn't change, but maybe it does, we don't have enough information

5. There are four versions (*"isotopes"*) of lead metal atoms. Their masses (units are *amu*) are listed below, along with how much of all lead is each type.

Type of Lead	Mass (amu)	% Abundance
A	203.97	?
B	205.97	24%
C	206.98	20%
D	207.98	54%

a) What number goes where "?" is ?

It is 100-(24+20+54)=2%

b) What is the average mass of lead?

$$\frac{(0.02)(203.97)+(0.24)(205.97)+(0.20)(206.98)+(0.54)(207.98)}{0.02+0.24+0.20+0.54}=207.22\,amu$$

6. To the right, are the four sections one which you were marked for your *Economics 203: Why Bitcoin is the Future* course.

Section	Contribution to Your Mark	You Scored...
Knowledge	40%	88%
Inquiry	20%	80%
Communication	10%	94%
Application	30%	84%

Calculate your mark for the course.

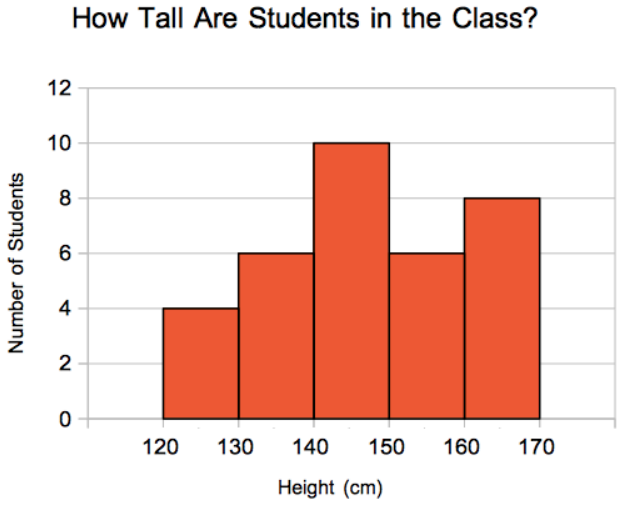
$$(88)(0.4)+(80)(0.2)+(94)(0.1)+(84)(0.3)$$
$$=85.8=86\%$$

Frequency Tables and Histograms

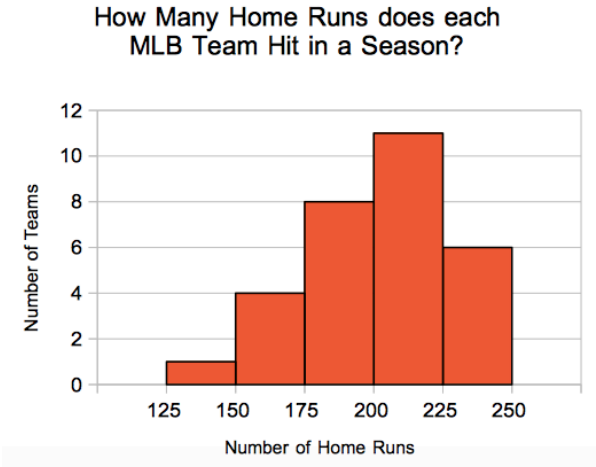
7. The following is the height data for people in your class.
Tally the values, and find the frequency for each interval as needed for the table.
Then, create a histogram for data.
Make sure to include a title and label your axes.

121	140	134	155	127	143	138
159	162	148	153	164	131	145
162	140	149	123	142	157	129
131	149	169	160	155	145	137
160	151	143	164	137	164	

Height (cm)	Tally	Number of Students
$120 \leq x < 130$	𝍸	4
$130 \leq x < 140$	𝍸 |	6
$140 \leq x < 150$	𝍸 𝍸	10
$150 \leq x < 160$	𝍸 |	6
$160 \leq x < 170$	𝍸 |||	8



8. Convert this histogram to an interval table.



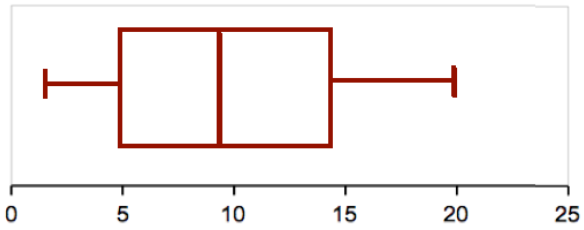
Home Runs Hit	Number of Teams
$125 \leq x < 150$	1
$150 \leq x < 175$	4
$175 \leq x < 200$	8
$200 \leq x < 225$	11
$225 \leq x < 250$	6

Quartiles, IQR and Box-and-Whisker Plots

9. Determine Q1, Q2 and Q3 for the following data.
Then, calculate the IQR.
Then, draw a box-and-whisker plot for it.

2
2
4
6 ← Q1 = (4+6)/2 = 5
6
8
10 ← Q2 = (8+10)/2 = 9
10
14
14 ← Q3 = (14+14)/2 = 14
16
20

IQR = Q3-Q1 = 14-5 = 9

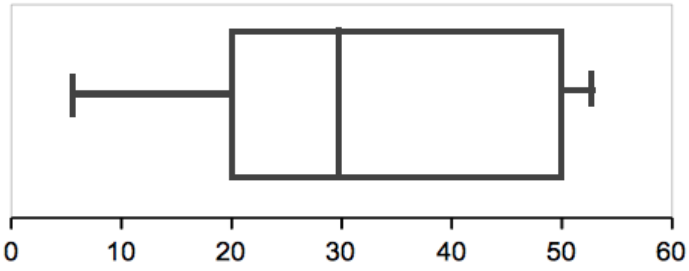


10. Given this box-and-whisker plot, state Q1, Q3, the IQR and the median.

Median = Q2 = 30

Q1 = 20

Q3 = 50



IQR
= Q3 – Q1
= 50 – 20
= 30

11. Here are the number of questions answered correctly by 14 students on the *Chem 13 News Exam*, which has 40 questions.

16 20 22 26 26 27 29 32 34 36 36 37 38 39

a) What percentile is the student who scored 32 ?

$$p = 100\left(\frac{L+0.5\,E}{n}\right) = 100\left(\frac{7+0.5\,E}{14}\right) = 53.6$$

b) What score would a student have needed to be in the 80th percentile?

$$R = \frac{80}{100}(14+1) = 12^{th}\ place$$ --> Requires a score of 37.

Variance and Standard Deviation

12. Calculate the standard deviation, and variance, of each set of data. Assume that each set of data is the *population*.

a) 2, 2, 4, 6, 10, 36

x	x-μ	$(x-\mu)^2$
2	2-10=-8	$(-8)^2=64$
2	2-10=-8	$(-8)^2=64$
4	4-10=-6	$(-6)^2=36$
6	6-10=-4	$(-4)^2=16$
10	10-10=0	$(0)^2=0$
36	36-10=26	$(26)^2=676$
Sum:		**856**

$$\sigma=\sqrt{\frac{856}{6}}=\sqrt{142.67}=11.94$$

Variance is $\sigma^2 = 142.67$

b) 2, 2, 4, 6, 8, 8

x	x-μ	$(x-\mu)^2$
2	2-5=-3	$(-3)^2=9$
2	2-5=-3	$(-3)^2=9$
4	4-5=-1	$(-1)^2=1$
6	6-5=1	$(1)^2=1$
8	8-5=3	$(3)^2=9$
8	8-5=3	$(3)^2=9$
Sum:		**38**

$$\sigma=\sqrt{\frac{38}{6}}=\sqrt{6.33}=2.52$$

Variance is $\sigma^2 = 6.33$

c) The data from (a) above has a range of 34, and the data from (b) has a range of 6. Does this relate to the standard deviation? Explain.

- A wider spread of data = a higher standard deviation.
- Standard deviation measures how far, on average, data is *away* from the mean.
- If there's more data that's farther away, there will be a higher standard deviation.

d) Recalculate the standard deviation of each data set, assuming that they are *samples* instead of populations. Note: This is **not** a long question – it is simply plugging in numbers you already know.

- The calculations would be exactly as found in (a) and (b), except $\bar{x}$ is written in place of μ.

For (a), $s=\sqrt{\frac{856}{5}}=\sqrt{171.2}=13.08$

For (b), $s=\sqrt{\frac{38}{5}}=\sqrt{7.6}=2.76$

These standard deviations are slightly higher than if we assume we have a *population* because a sample standard deviation must account for some unknown, uncollected data which may be farther from the mean than the data we did collect.

z-scores

13. The data {2, 2, 4, 6, 10, 36} has mean 10 and standard deviation of 11.94.

a) What is the z-score of each data point?

x=2: $z=\frac{x-\mu}{\sigma}=\frac{2-10}{11.94}=-0.67$

x=4: $z=\frac{x-\mu}{\sigma}=\frac{4-10}{11.94}=-0.50$

x=6: $z=\frac{x-\mu}{\sigma}=\frac{6-10}{11.94}=-0.34$ <-- data below the mean should have negative *z*

x=10: $z=\frac{x-\mu}{\sigma}=\frac{10-10}{11.94}=0$ <-- Data that EQUALS the mean has *z*=0

x=36: $z=\frac{x-\mu}{\sigma}=\frac{36-10}{11.94}=+2.18$ <-- data above the mean should have positive *z* and the farther away the number is, the bigger it should be.

b) If a 7th data point was added, with a z-score of +1.09, what would its value have been?

$$1.09=\frac{x-10}{11.94}$$

$$x=1.09(11.94)+10=23.01\approx 23$$

14. For a different set of data, which you don't have access to, a data point with x=12 has z=-1.4 and a data point with x=16 has z=-0.4. What is the mean and standard deviation of the data?

^^ Note. Even before we start, we can see that both 12 and 16 have negative *z*, so they are both **below** the mean.

You'll have to solve a system of equations for this one.

$-1.4=\frac{12-\mu}{\sigma}$ $\quad$ $-0.4=\frac{16-\mu}{\sigma}$

$-1.4\sigma=12-\mu$ $\quad$ $-0.4\sigma=16-\mu$

I choose elimination:

$-1.4\sigma=12-\mu$

$-0.4\sigma=16-\mu$

----------(subtract)

$-1\sigma=-4$

$\sigma=4$

and so $\mu=16+0.4\sigma=16+0.4(4)=16+1.6=17.6$