Pulse rates of males (2.3.11) uses ex (2.2.21)

60, 74, 86, 54, 90, 80, 66, 68, 68, 56, 80, 62, 74, 60, 52, 60, 66, 64, 64, 46, 68, 58, 68, 70, 56, 66, 78, 68, 62, 70, 72, 74, 64, 50, 70, 58, 60, 88, 84, 76

Begin with lower class limit of 40 and use a class width of 10.

range = 90−46 = $\frac{\text{max val}}{-\text{min val}}$

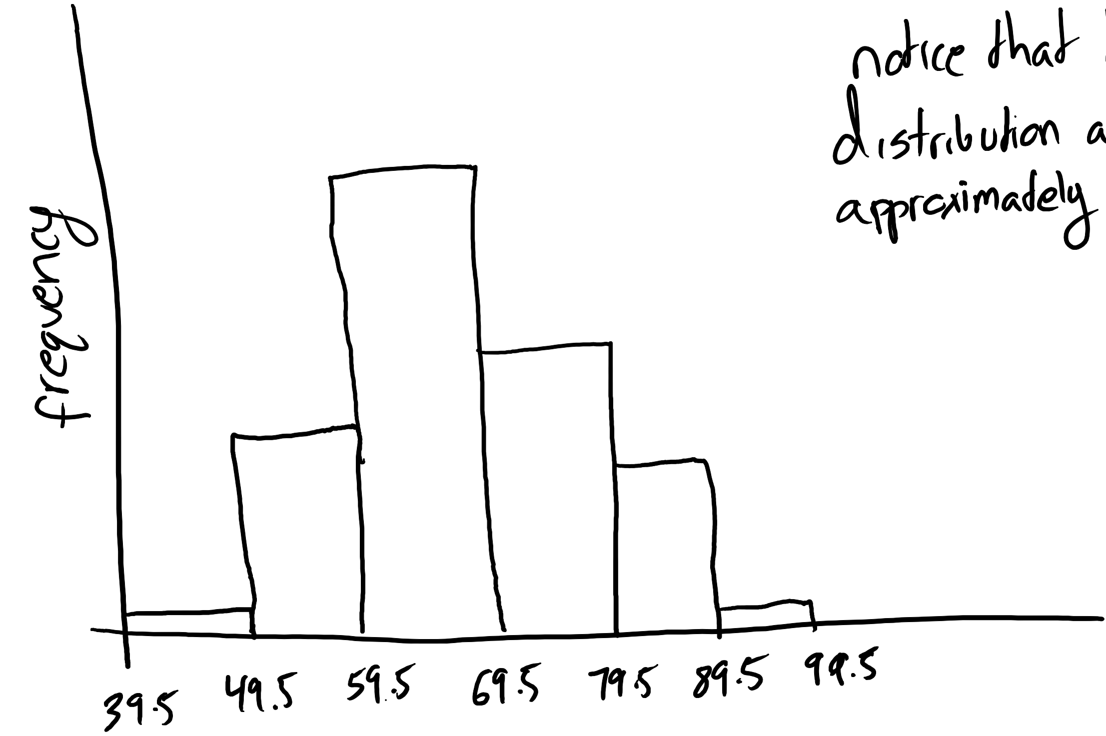
	range of class
class 1:	40 - 49
class 2:	50 - 59
class 3:	60 - 69
class 4:	70 - 79
class 5:	80 - 89
class 6:	90 - 99

classes based on range of 46-90 class width 10 and lower class limit 40

Next, count how many of the numbers in the data set fall inside each class.

class #	frequency
1	1
2	7
3	17
4	9
5	5
6	1

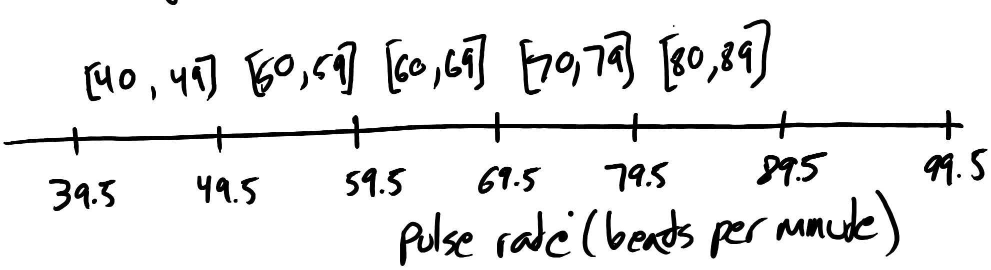
Based on the data, we make the following histogram:



notice that the distribution appears approximately normal.

Label y-axis with freq count, usually from zero.

We label the x-axis with the boundaries of the classes:



Let's look now at another example:

Ex) (2.3.14) uses data set specified in (2.2.24) earthquake depths Data set 16, appendix B. Begin with a lower class limit of 1.00km and use a class width of 4.00km. See data on next page.

The following classes are chosen

	equake depth (km)
class 1 :	1.00 – 4.99
class 2 :	5.00 – 8.99
class 3 :	9.00 – 12.99
class 4 :	13.00 – 16.99
class 5 :	17.00 – 20.99

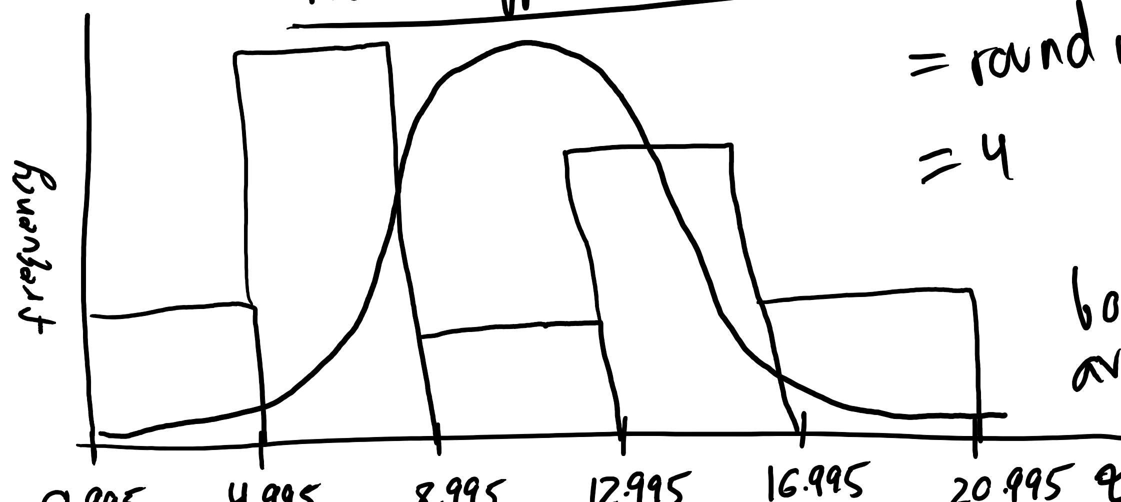
depth (cm)	frequency	boundaries
1.00 - 4.99	7	0.995, 4.995
5.00 - 8.99	21	4.995, 8.995
9.00 - 12.99	4	8.995, 12.995
13.00 - 16.99	12	12.995, 16.995
17.00 - 20.99	6	16.995, 20.995

notice class width is rounded up value of upper class limit minus lower class limit

= round up (3.99)

= 4

The following histogram then results: not approx normal



x-axis has class boundaries to avoid gaps.

This histogram is not bell shaped ; distribution does not appear normal.

range = max val - min val
= 18.9 - 2.0 = 16.9

Here is the data:

6.6	2.2	18.5	7.0	13.7	5.9	5.3	5.9	4.7	14.5
2.0	14.8	8.1	18.6	4.5	17.7	15.9	15.1	8.6	5.2
15.3	5.6	10.0	8.2	8.3	9.9	13.7	8.5	8.2	7.9
17.2	6.1	13.7	5.7	6.0	17.3	4.2	14.7	15.2	3.3
3.2	9.1	8.0	18.9	14.2	5.1	5.7	16.4	10.1	6.4

range 2.0 to 18.9 | told to start at 1.00 with class width of 4.

Summary so far

Chapter 1: population vs sample, good sampling methods, parameter and statistic

Chapter 2: frequency distributions & histograms, describing the distribution (is it approx. normal?), skewness, outliers.

next: quantitative measures of data

Chapter 3 (this and next lectures): summarize and describe the important characteristics of a data set with descriptive statistical quantities (mean, median, mode, std. deviation)

Suppose $X = \{x_1, \ldots, x_n\}$ a set of n observations. X is a set

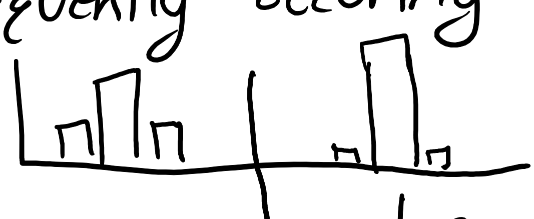
Then: range $= \max(X) - \min(X) \geq 0$

mean $= \mu = \dfrac{\sum X}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ (often dangerous when there are outliers)

(mu)

sum of observations over total # observations.

median: middle of the sorted data (arranged in increasing or decreasing magnitude) mode : the most frequently occuring value in the data set.



The variance is a measure of value variation in a data set $\bar{x} = \text{mean}(x)$

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

std dev= sqrt(variance)

Ex)

X (score)	$x-\bar{x}$ (deviation)	$(x-\bar{x})^2$
85	-5	25
90	0	0
95	5	25
88	-2	4
92	2	4

$$\bar{x} = \frac{85+90+95+88+92}{5} = \frac{450}{5} = 90 = \left(\frac{\sum x}{n}\right)$$

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

$$\sum(x-\bar{x})^2 = 25+25+8 = 58$$

notice: $\sum(x-\bar{x}) = 0$ does <u>not</u> imply $\sum(x-\bar{x})^2 = 0$!

$$s^2 = \frac{58}{5-1} = \frac{58}{4} = 14.5 = \frac{\sum(x-\bar{x})^2}{n-1}$$

standard deviation is the square root of the variance: std dev for above example is $s = \sqrt{14.5} \approx 3.81$ (note $s \geq 0$).

note: R (and related software) can easily be used to calculate these statistical quantities for large data sets.

using the other (equivalent) formula:

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{85^2 + 90^2 + 95^2 + 88^2 + 92^2 - \frac{(\sum x)^2}{n}}{n-1}$$

Also, $\sum x = 450 \Rightarrow (\sum x)^2 = 450^2 = 40500$

$$s^2 = \frac{40558 - 40500}{4} = \frac{58}{4} = 14.5$$

Ex) suppose $X = \{33, 2, 8, 2, 20\}$

$\text{sort}(x) = \{2, 2, 8, 20, 33\}$

median $= 8$; mode $= 2$

$$\text{mean} = \frac{2+2+8+20+33}{5} = \frac{65}{5} = 13$$