Chapter 3

population = all the elements

sample = only some of the elements

chapter 2 ⇒ graphical & tabular summarizing

chapter 3 ⇒ Numerical Measures for summarizing Data

sample statistics ⇒ Numerical measures computed from sample data

population parameters ⇒ Numerical measures computed from pop. data

"In statistical inference, a sample statistic is refered to as the point estimator of the corresponding population parameter." p. 99

Measures of location (get one value to represent all)

sample mean $= \bar{x} = \dfrac{\sum x_i}{n}$

$\bar{x} =$ "x bar"

$x_i =$ particular value

$n =$ # of observations in sample

example

wage set 1 $\begin{cases} x_1 = 12, \; x_2 = 15, \; x_3 = 19 \\ x_4 = 13, \; x_5 = 11 \end{cases}$

$$\bar{x}_1 = \frac{12 + 15 + 19 + 13 + 11}{5} = 14$$

wage set 2

$x_1 = 8, \; x_2 = 15, \; x_3 = 22$

$x_4 = 7 \quad x_5 = 18$

$$\bar{X}_2 = \frac{8 + 15 + 22 + 7 + 18}{5} = 14$$

Chart 1

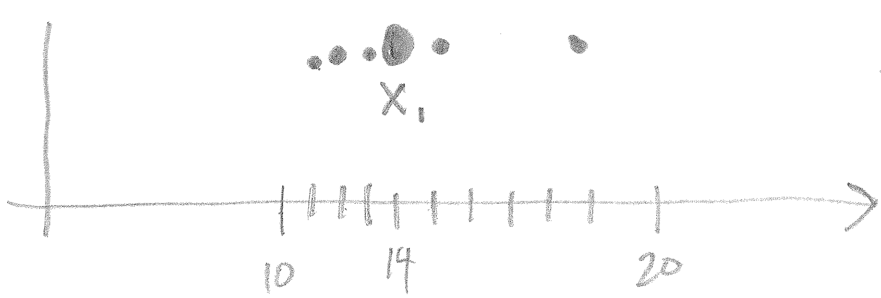
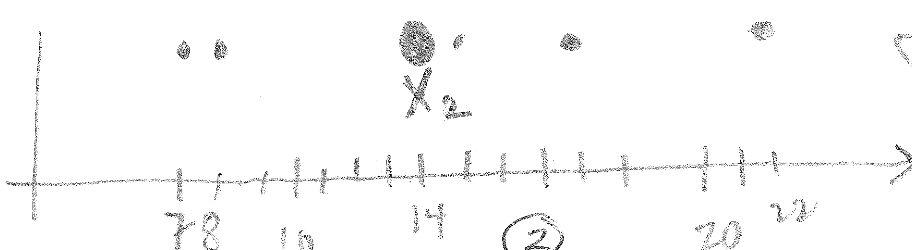


chart 2



more spread

Measures of variability (How reliable is mean)

sample standard deviation $= s = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n-1}}$

deviation from $\bar{x}$

Sort of like average of all deviations; it tells you how well the mean represents its data points

wage set 1

$\bar{x}_1 = 14$

$n = 5$

$n-1 = 5-1 = 4$

x	$x - \bar{x}$	$(x-\bar{x})^2$
12	$12-14 = -2$	$-2^2 = 4$
15	$15-14 = 1$	$1^2 = 1$
19	$19-14 = 5$	$5^2 = 25$
13	$13-14 = -1$	$-1^2 = 1$
11	$11-14 = -3$	$-3^2 = 9$
	$\underbrace{\Sigma(x-\bar{x}) = 0}_{\text{always!}}$	$\Sigma(x-\bar{x})^2 = 40$

that is why we square

$s = \sqrt{\dfrac{40}{4}} =$

$s = \sqrt{10}$

$s_{\text{wage set 1}} = 3.162278$

wage set 2

$\bar{x}_2 = 14$
$n = 5$
$n-1 = 5-1 = 4$

x	$x - \bar{x}$	$(x-\bar{x})^2$
8	$8-14=-6$	$-6^2 = 36$
15	$15-14 = 1$	$1^2 = 1$
22	$22-14 = 8$	$8^2 = 64$
7	$7-14 = -7$	$-7^2 = 49$
18	$18-14 = 4$	$4^2 = 16$
	$\underbrace{\Sigma(x-\bar{x}) = 0}_{\text{Always!}}$	166

$$s_{\text{wage set 2}} = \sqrt{\frac{166}{4}} = 6.442049$$

---

which mean represents its data points more fairly?

$s_1 = 3.16$

$s_2 = 6.44$ "more spread"

Homework #42 ch. 3

5-Number Summary

Excel

Min = 30

Q1 = 50.5

Q2 = 66

Q3 = 87.75

Max = 208 (yankees)

Interquartile Range $\left.\right\}$ 87.75 − 50.5 = 37.25

Lower Limit $\left.\right\}$ = 50.5 − 1.5 * 37.25 = 0 cuz value below zero

Upper Limit $\left.\right\}$ 87.75 + 1.5 * 37.25 = 143.625

Smallest inside Lower Limit $\left.\right\}$ = 30

Biggest = 124

By Hand

Min = 30

Count = 8

Q1 i = .25 * 8 = 7.5 ⇒ 8 ⇓ 49

Q2 = 66

Q3 i = .75 * 8 = 22.5 ⇒ 23 ⇓ 88

Max = 208

Interquartile Range $\left.\right\}$ 88 − 49 = 39

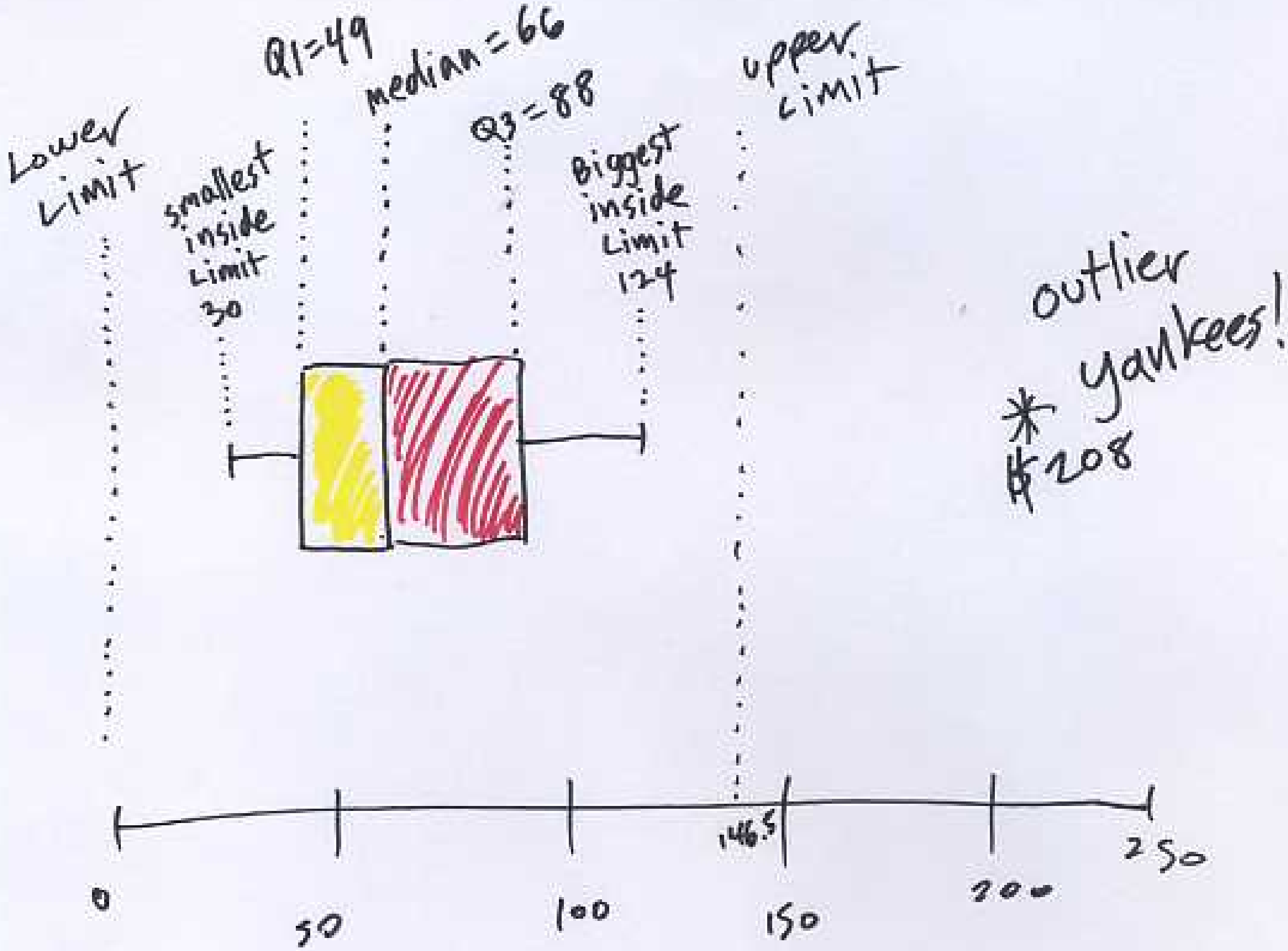
Lower Limit $\left.\right\}$ = 49 − 39 * 1.5 = 0 cuz value below zero

Upper Limit $\left.\right\}$ 88 + 1.5 * 39 = 146.5

Small = 30

Biggest inside upper Limit $\left.\right\}$ = 124

Box plot →



By Hand

Box plot

See Excel sheet # 42 for Excel version