Panel 1

Last Time:

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

sort

Lower Quartile: Q_1 : compute $L_1 = 0.25 \cdot N$

Upper Quartile: Q_3 : $L_3 = 0.75 N$

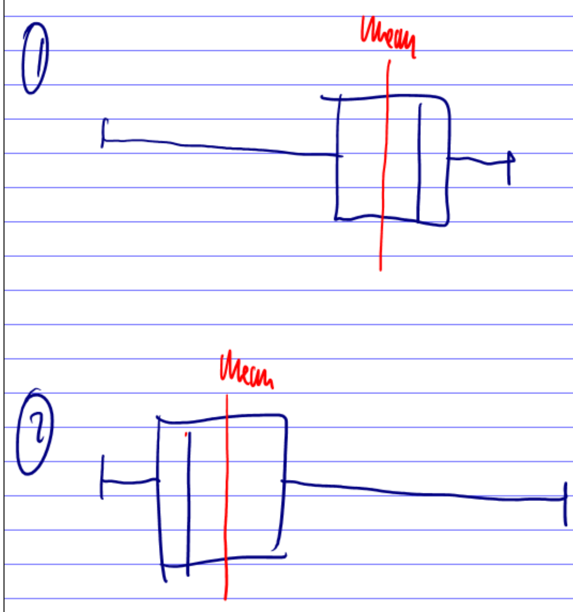
IQR: Inter Quartile Range $Q_3 - Q_1$

Box Plot: 5-Number Summary.

n-th Percentile: $L_n = n^{th} \cdot N$

1

Panel 2



You can estimate mean / and std. dev. from a box plot

2

Panel 3

Estimating the Standard Deviation

s is difficult to compute, so want to estimate it

Empirical Rule: In a bell-shaped distribution we have

- 68% of data values are between $\bar{x} - s$ and $\bar{x} + s$
- 95% of data values are between $\bar{x} - 2s$ and $\bar{x} + 2s$
- 100% between $\bar{x} - 3s$ and $\bar{x} + 3s$

$\Rightarrow$ Range $\simeq 4s$ $\Rightarrow$ $s \simeq \frac{\text{Range}}{4}$

3

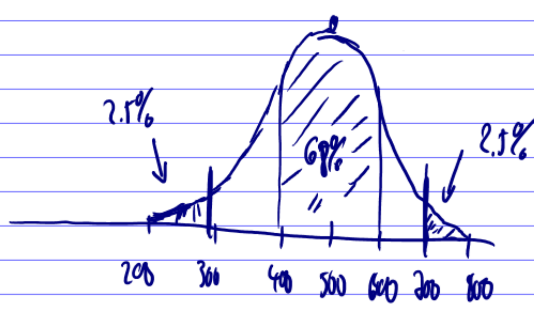
Panel 4

Ex: SAT scores have $\bar{x} = 500$ and $s = 100$

$\Rightarrow$ 68% of SAT scores are between 400 and 600

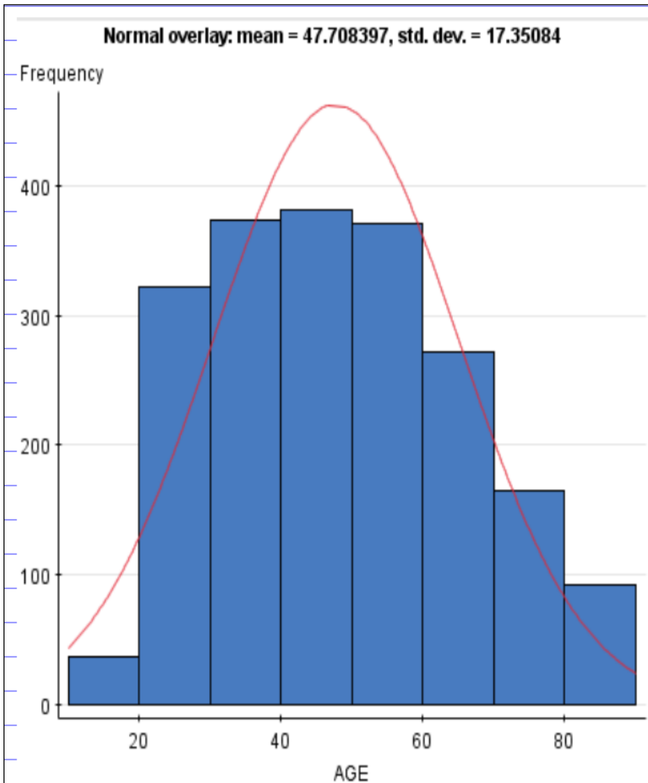
95% of SAT scores are between 300 and 700

100% of SAT scores are between 200 and 800



4

Panel 5



min = 18

max = 89

Q_1 = 34

Q_2 = 47

Q_3 = 60

$s \approx \frac{\text{Range}}{4} = \frac{89-18}{4} = \frac{71}{4} = 17.75$

IQR = 60-34 = 26

$s \approx \frac{3}{4} \cdot 26 = 19.5$

5

Panel 6

IQR and Standard Deviation

Recall that mean is impacted by outliers, (median is not)

Std. dev. is also impacted by outliers (Q1 and Q3 are not)

Theorem IQR $\approx \frac{4}{3} s$ for bell-shaped (normal) distributions

$\Rightarrow s \approx \frac{3}{4}$ IQR

6

Panel 7

Outliers: are particularly large or small data values.

Def. An outlier is a data value that is more than $1.5 \cdot IQR$ larger than Q_3, or more than $1.5 \cdot IQR$ less than Q_1

$4 + 4.5$

$4 \quad 1.5 \cdot 3$

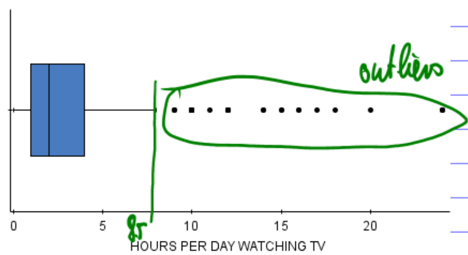
$Q_1 = 1$

$Q_3 = 4$

$IQR = 3$ so outliers are: bigger than $8.5 = Q_3 + 1.5\, IQR$

less than $-3.5 = Q_1 - 1.5\, IQR$

$1 - 4.5$



$1.5 \cdot 3 = 4.5$

7

Panel 8

Exam 1:

28, 32, 33, 28, 29, 30, 30, 28, 32, 30

Find everything: 28, 28, 28, 29, 30, 30, 30, 32, 32, 33 $\quad N = 10$

x	x^2
28	784
28	784
28	784
29	841
30	900
30	900
30	900
32	1024
32	1024
33	1089
300	9030

$L_1 = 0.25 \cdot 10 = 2.5 \Rightarrow Q_1 = 29$

$L_2 = 0.5 \cdot 10 = 5$, median 30

$L_3 = 0.75 \cdot 10 = 7.5 \Rightarrow Q_3 = 32$

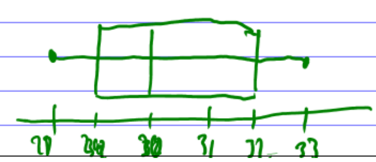
$\bar{x} = \frac{300}{10} = 30$

$s^2 = \frac{1}{9}\left(9030 - \frac{300^2}{10}\right) = \frac{30}{9} = 3.33$, $s = 1.824 = \sqrt{3.33}$

check: $s \approx \frac{\text{Range}}{4} = \frac{5}{4} = 1.25$

$IQR \cdot \frac{3}{4} = 3 \cdot \frac{3}{4} = \frac{9}{4} = 2.25$

mode: 28 and 30



8