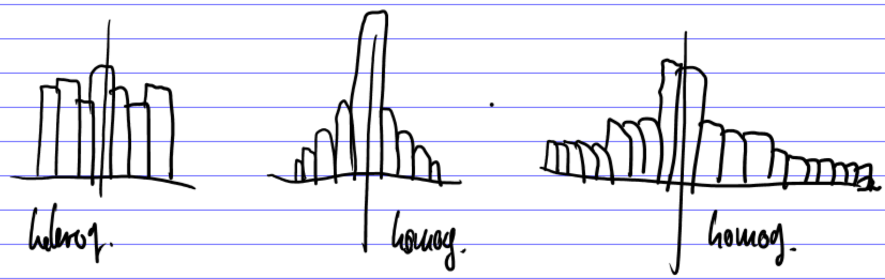
Panel 1

Last time:

Exam I on Monday

homogeneous / heterogeneous distributions



Variance

s^2 or σ^2 : $\frac{1}{n-1}\sum (x-\bar{x})^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$

Standard Deviation : is square-root of variance.

s or σ

1

Panel 2

Data: 3, 4, 7, 1, 9, 8 Find std. dev. s.

s^2:

x	x^2
3	9
4	16
7	49
1	1
9	81
8	64
32	220

$\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) =$

$= \frac{1}{5}\left(220 - \frac{32^2}{6}\right) =$

$= \frac{1}{5}(49.33) = 9.867$ is variance.

Thus: std. dev. $s = \sqrt{9.867} = 3.141$

2

Panel 3

More numeric data descriptors

Quartiles: 2 quartiles are defined as follow

Lower Quartile (Q_1): that number such that 25% are less, 75% are bigger

Q_2: already called median

Upper Quartile (Q_3): that number such that 75% are less, 25% are bigger

Ex: Data 1, 2, 3, 4, 5, 6, 7 - Find quartiles

1st, split data in half: [1, 2, 3] 4 [5, 6, 7] (4 is median)

2nd split each half: [1] 2 [3] 4 [5] 6 [7]. $Q_1 = 2$, $Q_3 = 6$

3

Panel 4

How to find Quartiles

1. Sort data, starting with smallest
2. Find N (sample size)
3. Compute position $L_1 = 0.25 \cdot N$
   - if L_1 is whole, pick # between L_1 and next position
   - else pick value at next position $\Rightarrow Q_1$
4. For Q_3, use $L_3 = 0.75 \cdot N$
5. For Q_2, or median use $L_2 = 0.5 \cdot N$

4

Panel 5

Ex: Data 1, 2, 3, 4, 5 sorted, $N = 5$

$L_1 = 0.25 \cdot 5 = 1.25 \Rightarrow Q_1 = (2^{nd}\ \#) = 2$

$L_3 = 0.75 \cdot 5 = 3.75 \Rightarrow Q_3 = (4^{th}\ \#) = 4$

$L_2 = 0.5 \cdot 5 = 2.5 \Rightarrow Q_2 = (3^{rd}\ \#) = 3$

Ex: Cotinine level of 11 smokers is as follows:

0, 87, 173, 253, 1, 103, 173, 265, 1, 112, 198

Find Q_1, Q_3, Median: 0, 1, 1, 87, 103, 112, 173, 173, 198, 253, 265

$Q_1: L_1 = 0.25 \cdot N = 2.75 \Rightarrow Q_1 = 1$ | $Q_2 = 6^{th}\ \# = 112$ | $Q_3: L_3 = 8.25 \Rightarrow Q_3 = 198$

Panel 6

Recall: Your SAT score is in the 95th percentile, i.e. 95% of all score are worse that yours, 5% are better.

Def: x^{th} percentile is that number, for which $x\%$ are less, $(100-x)\%$ are more than it.

You compute it by figuring out the $x \cdot N^{th}$ position

Ex: 85^{th} percentile of previous data:

$L_x = x \cdot N = 0.85 \cdot 11 = 9.35 \Rightarrow$

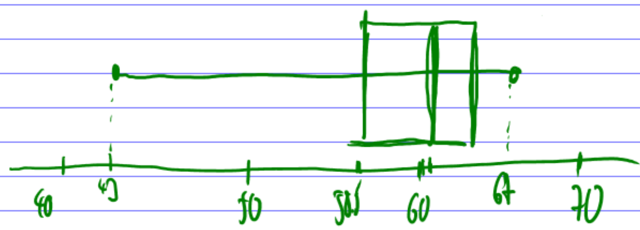
85th percentile is: 253 (10^{th} number)

Panel 7

Have a variety of statistics that describe a numeric variable. Want to combine them into a convenient picture: **Box Plot**

43	51	53	55	57	58	58	59	60	61
61	61	61	61	62	63	64	64	65	65
65	66	66	66	66	66	66	67		

1.) Draw a line from min to max, horizontally



shifted to left, i.e. mean < median (61.7)

2.) Draw vertical lines at Q_1, Q_3, and median

3) Complete the box.

7