Section 3.1: Measures of Central Tendency

Friday, May 03, 2013
1:48 PM

numerical measures - another way to summarize a data set

(graphs → visual summary)

- particularly of use when comparing two data sets whose values look very similar

so numerical measures are quantities that describe a data set

→ called **parameters** for population

→ called **statistics** for samples

we'll look at 3 types:

- measures of centre
- measures of dispersion (spread or variability)
- measures of relative standing

---

measures of centre:

mean (average): add all the values, and then divide by the number of data points

population mean

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

mean $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$

Greek letter "mu"

where x_i is the value of the i^{th} data point

n is the number of points

---

sample mean $\bar{x}$ = same calculation

one problem with the mean - it's greatly affected by outliers

data points with values very different from the majority of points

---

median - if you write your data in the form of an ordered list, it's the middle value

- if you have an even number of points, it's the average of the two middle points

---

example: (fictitious data!)

starting salary: $90 000

$50,000
$24,000
$65,000
$3,000,000

mean = $645,800 ← a bit misleading
median = $65,000

---

note: if there are lots of data points, say 75, how do you determine which is the middle point?

$$\text{position} = \tfrac{1}{2}(n+1)$$

the median will be the 38th point

for 76 points,

position = 38.5

↑ so take the average of the 38th and 39th points

What happens to the mean when you multiply all data points by 5?

-multiplied by 5 also

---

weighted mean:

x	f
1	5
2	4

} frequency table

1	5
2	4
3	2

frequency table

↑ number of times x appears in the data set

$$\bar{x} = \frac{\sum x_i f_i}{n}$$

$$= \frac{5 \cdot 1 + 4 \cdot 2 + 2 \cdot 3}{11}$$