9/14/09

Sample variance

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$$

} decimal places 2 extra

Variance has (units)^2

measures typical squared deviation from the mean.

Sample std. dev.

$$s = \sqrt{s^2} = \sqrt{\text{sample variance}}$$

} 1 extra decimal place

measures typical deviation from the mean

(i.e. $\bar{x} = 12$, $s = 9.4$ means 12 ± 9.4 is typical data)

Properties

$s^2 \geq 0$ (variance never negative)

$s \geq 0$

if $s = 0$ then there is NO variation in the data → all observations are the same.

---

§1.6 Grouped Data:

9. From a certain population of the freshwater sculpin, *Cottus rotheus*, a sample of 100 fish were dissected to determine the number of tail vertebrae in each. The data collected are summarized below. Find the mean, median, and standard deviation for this data set.

frequency = count

$\tilde{X} = 21$

No. of vertebrae (X_i)	f_i (frequency)	$f_i X_i$	$f_i X_i^2$
20	3	60	1,200
21	51	1,071	22,491
22	40	880	19,360
23	6	138	3,174
	$n =$ 100	2,149	46,225

frequency table

$$\bar{X} = \frac{\sum f_i X_i}{n} = \frac{3\cdot 20 + 51\cdot 21 + 40\cdot 22 + 6\cdot 23}{100}$$

$$= \frac{2149}{100} = 21.49 \approx 21.5 \text{ vertebrae}$$

Sum over all the categories

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

$$= \frac{\left(3\cdot 20^2 + 51\cdot 21^2 + 40\cdot 22^2 + 6\cdot 23^2\right) - \frac{(2149)^2}{100}}{99}$$

$$= \frac{46225 - \frac{(2149)^2}{100}}{99} = .43 \text{ (vertebrae)}^2$$

(typical squared deviation)

Std. dev. $s = \sqrt{.43} = .7$ vertebrae

typical: $\bar{X} \pm s = 21.5 \pm .7$ vertebrae

HW: Ch. 1 : 1-7, 10, 14