Section 5.5 The Variance and Standard Deviation

Math 141

Main ideas

Mean and **variance** of data/outcomes $x_1, x_2, \ldots, x_r$ with *frequencies* $f_1, f_2, \ldots, f_r$:

Population

$$\mu = \frac{x_1 f_1 + x_2 f_2 + \cdots x_r f_r}{N} = x_1\left(\frac{f_1}{N}\right) + x_2\left(\frac{f_2}{N}\right) + \cdots + x_r\left(\frac{f_r}{N}\right)$$

$$\sigma^2 = \frac{(x_1-\mu)^2 f_1 + (x_2-\mu)^2 f_2 + \cdots + (x_n-\mu)^2 f_r}{N} = (x_1-\mu)^2\left(\frac{f_1}{N}\right) + (x_2-\mu)^2\left(\frac{f_2}{N}\right) + \cdots + (x_r-\mu)^2\left(\frac{f_r}{N}\right)$$

Sample (used for estimating corresponding values for the population)

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots x_r f_r}{n} = x_1\left(\frac{f_1}{n}\right) + x_2\left(\frac{f_2}{n}\right) + \cdots + x_r\left(\frac{f_r}{n}\right)$$

$$s^2 = \frac{(x_1-\bar{x})^2 f_1 + (x_2-\bar{x})^2 f_2 + \cdots + (x_n-\bar{x})^2 f_r}{n-1}$$ ($n-1$ makes s^2 and s better estimates for σ^2 and σ)

Mean and **variance** of data/outcomes $x_1, x_2, \ldots, x_r$ with *probabilities* $p_1, p_2, \ldots, p_r$:

$$\mu = x_1 p_1 + x_2 p_2 + \cdots + x_r p_r$$

$$\sigma^2 = (x_1-\mu)^2 p_1 + (x_2-\mu)^2 p_2 + \cdots + (x_r-\mu)^2 p_r$$

Mean and **variance** for n binomial trials with *probability of success* p:

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n$$

$$\sigma^2 = np(1-p)$$

In all cases, **standard deviation** is the square root of the variance.

Chebychev's Inequality (page 275): the fraction of all values that are within k standard deviations of the mean is $1 - \left(\frac{1}{k}\right)^2$.

Optional: alternate formula for variance (top of page 272).

Problems

1. Suppose you got an 85 on an exam. How good is your score of 85 if it came from the scores

   75, 78, 79, 80, 81, 82 and 85 — High score!

   65, 70, 75, 80, 85, 90 and 95 — Above average

2. Find the mean (average) of:

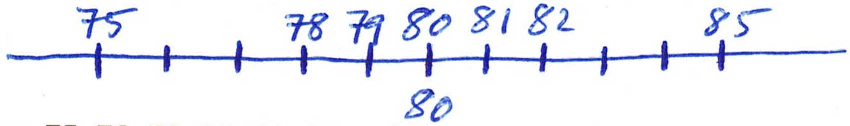
   75, 78, 79, 80, 81, 82 and 85 — 80

   65, 70, 75, 80, 85, 90 and 95 — 80

3. Which set of scores is more "spread out"? How can we measure/quantify "spread out"?

   75, 78, 79, 80, 81, 82 and 85 — Less spread out

   65, 70, 75, 80, 85, 90 and 95 — More spread out



4. For the scores 75, 78, 79, 80, 81, 82 and 85:
Find the deviation (distance, difference) of each score from the mean of 80:

$-5, -2, -1, 0, 1, 2, 5$

So the average deviation of each score from the mean of 80 is:

$$\frac{-5-2-1+0+1+2+5}{7} = 0$$

Find the average <u>absolute-valued</u> deviation of each score from 80:

$$\frac{5+2+1+0+1+2+5}{7} = \frac{16}{7} \approx 2.3$$

Find the average deviation <u>squared</u> of each score from 80 (the variance) and its square root:

$$\sigma^2 = \frac{(-5)^2+(-2)^2+\cdots+5^2}{7} \approx 8.57 \qquad \text{so } \sigma = 2.93$$

5. For the scores 65, 70, 75, 80, 85, 90 and 95 (which have mean 80), find the variance σ^2 and standard deviation σ:

$$\sigma^2 = \frac{(65-80)^2+\cdots+(95-80)^2}{7} = 100 \qquad \text{so } \sigma = 10$$

6. For the numbers 1, 2, 2, 3, 3, 3, 3, 3:

$$\mu = \frac{1\cdot 1 + 2\cdot 2 + 3\cdot 5}{8} = 1\left(\tfrac{1}{8}\right) + 2\left(\tfrac{2}{8}\right) + 3\left(\tfrac{5}{8}\right) = 2.5$$

$$\sigma^2 = \frac{(1-2.5)^2\cdot 1 + \cdots + (3-2.5)^2\cdot 5}{8} = (1-2.5)^2\left(\tfrac{1}{8}\right)+\cdots$$

$$\sigma = \sqrt{.5} \approx .71 \qquad = .5$$

7. Find variance and standard deviation for each investment. (Higher standard variance and standard deviation mean higher risk and unpredictability in investment.)

Let X_A = return for investment A

k	$\Pr(X_A = k)$
\$1000	.20
\$2000	.50
\$3000	.30

Let X_B = return for investment B

k	$\Pr(X_B = k)$
-\$1000	.30
\$ 0	.10
\$4000	.60

$E(X_A) = \$1000(.20) + \cdots + \$3000(.30) = \$2100$

$\sigma_A^2 = (1000-2100)^2(.20) + \cdots + (3000-2100)^2(.30) = 490\,000$

$\sigma_A = \sqrt{490\,000} = 700$

$E(X_B) = -\$1000(.30) + \cdots + \$4000(.60) = \$2100$

$\sigma_B^2 = (-1000-2100)^2(.30) + \cdots + (4000-2100)^2(.60) = 5\,490\,000$

$\sigma_B = \sqrt{5490000} \approx 2343$

8. Consider samples of students from two colleges. *We use the information from the samples as estimates for all of the students at each college.*

College A

Age	Students
18	1
19	2
20	4
21	2
22	1
23	0

$$\bar{x}_A = \frac{18 \cdot 1 + 19 \cdot 2 + \cdots + 22 \cdot 1}{10} = 20 \qquad \approx \mu_A$$

$$s_A^2 = \frac{(18-20)^2 \cdot 1 + \cdots + (22-20)^2 \cdot 1}{10-1} = \frac{4}{3} \qquad \approx \sigma_A^2$$

$$s_A = \sqrt{\frac{4}{3}} \approx 1.15 \qquad \approx \sigma_A$$

College B

Age	Students
18	3
19	1
20	2
21	2
22	1
23	1

$$\bar{x}_B = \frac{18 \cdot 3 + \cdots + 23 \cdot 1}{10} = 20 \qquad \approx \mu_B$$

$$s_B^2 = \frac{(18-20)^2 \cdot 3 + \cdots + (23-20)^2 \cdot 1}{10-1} = \frac{28}{9} \qquad \approx \sigma_B^2$$

$$s_B = \sqrt{\frac{28}{9}} \approx 1.76 \qquad \approx \sigma_B$$

9. Given several exam scores with mean 80 and standard deviation 5, using Chebyshev's Inequality, then the probability that a particular score is between:

70 and 90 is at least $1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4} = .75$

65 and 95 is at least $1 - \left(\frac{1}{3}\right)^2 = \frac{8}{9} \approx .89$

That is, *at least* 75% of the values are within 2 standard deviations of the mean.
That is, *at least* 89% of the values are within 3 standard deviations of the mean.

10. 70% free throw shooter. Shoot 2 shots. X = number of shots made.

k	Pr(X = k)
0	$C(2,0)(.70)^0(.30)^2 = .09$
1	$C(2,1)(.70)^1(.30)^1 = .42$
2	$C(2,2)(.70)^2(.30)^0 = .49$

$E(X) = 0\ (.09) + \cdots + 2\ (.49) = 1.4$

$\sigma^2 = (0-1.4)^2(.09) + \cdots + (2-1.4)^2(.49) = .42$

$\sigma = \sqrt{.42} \approx .65$

Notice that $E(X) = np = 2(.70) = 1.4$ and $\sigma^2 = np(1-p) = 2(.70)(.30) = .42$.

11. 70% free throw shooter. Shoot 3 shots. X = number of shots made.

k	Pr(X = k)
0	$C(3,0)(.70)^0(.30)^3 = .027$
1	$C(3,1)(.70)^1(.30)^2 = .189$
2	$C(3,2)(.70)^2(.30)^1 = .441$
3	$C(3,3)(.70)^3(.30)^0 = .343$

$E(X) = 0\ (.027) + \cdots + 3\ (.343) = 2.1$

$\sigma^2 = (0-2.1)^2(.027) + \cdots + (3-2.1)^2(.343) = .63$

$\sigma = \sqrt{.63} \approx .79$

Notice that $E(X) = np = 3(.70) = \underline{2.1}$ and $\sigma^2 = np(1-p) = 3(.70)(.30) = \underline{.63}$.