* Announcement :

① clobber policy

- Only final can clobber the midterm (one-directional)

② Midterm grading by the end of this week

③ HW7 : 10/12 ~ 10/19

④ Quiz 5 : Chapter 5.4 ~ 5.6

STAT 88: Lecture 20

After today's lecture

⇝ HW7 Q1, Q2

Contents

Warm up: Let the distribution of X be:

x	1	2	3
$(x-\mu_x)^2$	$(1.1)^2$	$(0.1)^2$	$(0.9)^2$
$P(X=x)$	0.2	0.5	0.3

(a) Find $\mu_X = E(X)$.

(b) Find the distribution of $(X - \mu_X)^2$ in table.

(c) Find $E((X - \mu)^2)$.

(a) $E(X) = 1 \times (0.2) + 2 \times (0.5) + 3 \times (0.3)$
$= 2.1$

(b) --

(c) $E((X-\mu)^2) = (1.1)^2 \cdot (0.2) + (0.1)^2 \cdot (0.5) + (0.9)^2 \cdot (0.3)$
$= 0.49$

//

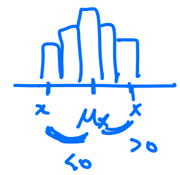
$Var(X)$

6.1. Variance and Standard Deviation

Expectation : center of a distribution
Standard deviation : average spread of a dist'n about the center

Variance Let X be a random variable and let $\mu_X = E(X)$. Define $D = X - \mu_X$, the deviation from the expected value. Note $E(D) = E(X - \mu_X) = 0$.

$= \underbrace{E(X)}_{= \mu_X} - \mu_X$



We define a measure called the variance of X by

$$\mathrm{Var}(X) = E(D^2) = E((X - \mu)^2).$$

cm² cm

We saw how to calculate this in the warm up. Note that the units of X are squared.

Standard deviation

$$\mathrm{SD}(X) = \sqrt{\underbrace{\mathrm{Var}(X)}_{cm^2}} = \sqrt{E((X - \mu)^2)}.$$

cm cm

Interpretation: "SD(X)" is roughly the "average" variation from the center.

Ex:

$E(Y) = (y - \mu_Y)^2$

$(y-\mu_Y)^2$	$(0.8)^2$	$(0.2)^2$	$(1.2)^2$
y	3	4	5
$P(Y = y)$	0.55	0.1	0.35

Calculate (1) $E(Y)$ (2) $\mathrm{Var}(Y)$ (3) $\mathrm{SD}(Y)$.

(1) $E(Y) = 3 \cdot (0.55) + 4 \cdot (0.1) + 5 \cdot (0.35)$
$= 3.8$

(2) $\mathrm{Var}(Y) = E((Y - \mu_Y)^2)$
$= (0.8)^2 \cdot (0.55) + (0.2)^2 \cdot (0.1) + (1.2)^2 \cdot (0.35)$
$= 0.86$

(3) $\mathrm{SD}(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{0.86}$

In Python:

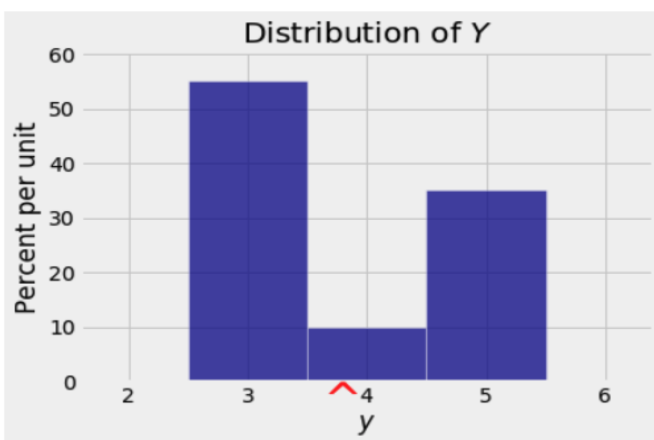
```
variance_table_Y
```

y	(y - E(Y))**2	P(Y = y)
3	0.64	* 0.55
4	0.04	* 0.1
5	1.44	* 0.35

```
var_Y = sum(variance_table_Y.column(1) * variance_table_Y.column(2))
sd_Y = var_Y ** 0.5   ← (Var_Y)^½
sd_Y
```

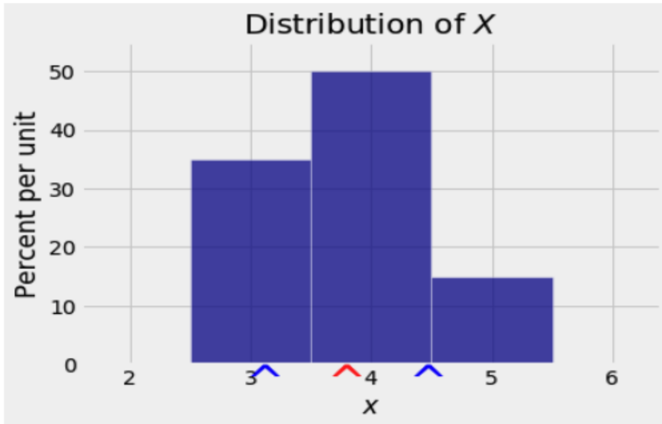
```
0.9273618495495703
```

Picture:



E(Y) = 3.8

Compare with



$E(X) = 3.8$

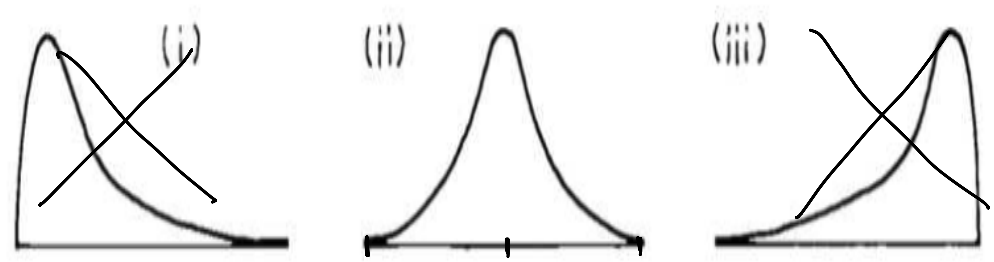
SD(X) vs SD(Y) ?

$<$

$SD(Y) > SD(X)$

Example: About 300 Stat 88 students at UC Berkeley, were asked how many college mathetmatics courses they had taken other than Stat 88. The average number of courses was about 1.1; the SD was about 1.5. ~~Would the histogram for the data look like (i), (ii), or (iii)?~~

X = # math courses



0 1.1 2

$x = 0$, $\quad (1.1)^2 = 1.21$, $\quad 1.1 = E(X)$

$$SD(X) = 1.5 = \sqrt{E((X-\mu)^2)}$$

$$Var(X) = E((X-\mu)^2)$$

$X = x \rightsquigarrow (X-\mu)^2 = (x-\mu)^2$

Min $x = 0$

Max $x = 2$

$$\text{Max } (x-\mu)^2 = (0-1.1)^2 = (1.1)^2 = 1.21$$

$\to E((X-\mu)^2) < 1.21$

$\to SD(X) = \sqrt{E((X-\mu)^2)} < \sqrt{1.21} < 1.5$

$\to SD(X) = 1.5$ not possible

6.2. Simplifying the Calculation

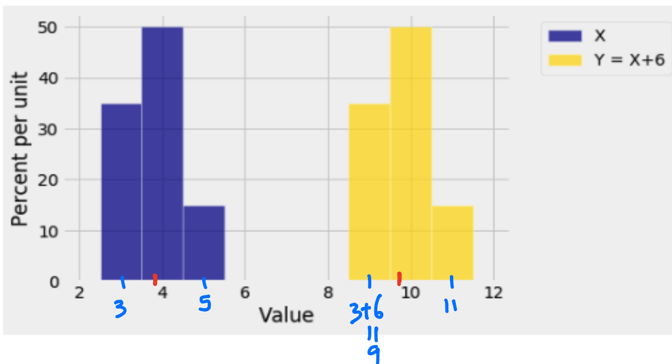
Linear Transformations

Celsius-Fahrenheit conversion:

$$Y = (9/5) \cdot X + 32.$$

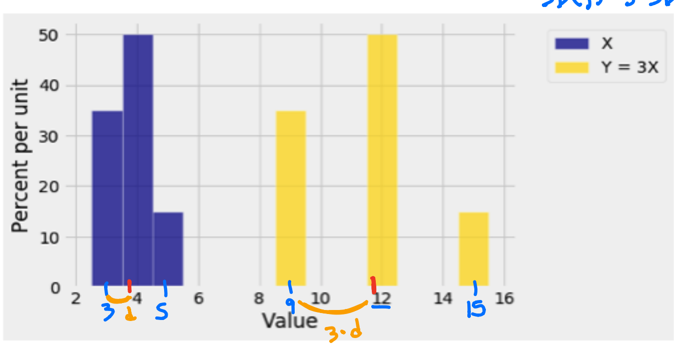
How does SD(Y) compare to SD(X)?

$E((X-\mu)^2)$



$SD(X+b) = SD(X)$

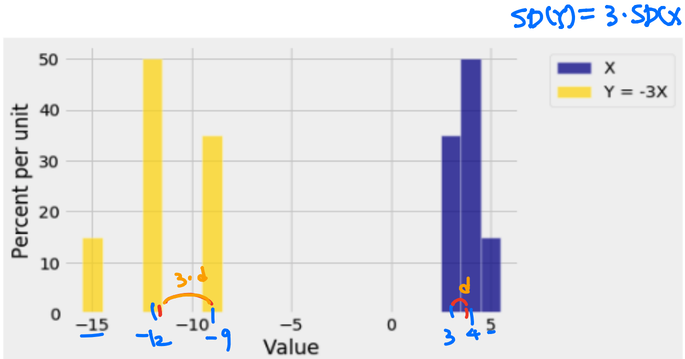
$SD(Y) = 3 \cdot SD(X)$



$a > 0$

$SD(aX) = a\, SD(X)$

$SD(Y) = 3 \cdot SD(X)$



$a < 0$,

$SD(aX) = |a|\, SD(X)$

So, we have

$$\mathrm{SD}(aX+b) = |a|\mathrm{SD}(X),$$

^2

and

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

Hence if $Y = (9/5)X + 32$, then

$$\mathrm{SD}(Y) = (9/5)\cdot\mathrm{SD}(X).$$

A Different Way of Calculating Variance An algebraic simplification for calculating variance:

= E(x) "constant"

*

$$\mathrm{Var}(X) = E((X-\mu_X)^2)$$

$$= E(X^2 - 2\mu_X X + \mu_X^2)$$

$$= E(X^2) - E(2\mu_X X) + E(\mu_X^2)$$

$$= E(X^2) - 2\cdot\mu_X \underbrace{E(X)}_{=\mu_X} + \mu_X^2$$

$$= E(X^2) - 2\mu_X^2 + \mu_X^2$$

$$= E(X^2) - \mu_X^2$$

$$= E(X^2) - (E(X))^2$$

Ex:

Y^2	9	16	25
y	3	4	5
$P(Y=y)$	0.55	0.1	0.35

Find $\text{Var}(Y) = E(Y^2) - E(Y)^2$.

$E(Y) = 3.8$

$E(Y^2) = 9 \cdot (0.55) + 16 \cdot (0.1) + 25 \cdot (0.35)$

$= 15.3$

$\text{Var}(Y) = 15.3 - (3.8)^2 = 0.86$

Example: (Exercise 6.5.5) Let $p \in (0,1)$ and let X be the number of spots showing on a flattened die that shows its six faces according to the following chances:

- $P(X=1) = P(X=6)$
- $P(X=2) = P(X=3) = P(X=4) = P(X=5)$
- $P(X = 1 \text{ or } X = 6) = p$ $\rightsquigarrow P(X=1)?$ $\overset{\shortparallel}{p/2}$

$\overset{\shortparallel}{P(X=1) + P(X=6)}$

$\rightsquigarrow P(X=1) = p/2$
$P(X=6) = p/2$

Find SD(X).

$\frac{7}{2}$ ✓

x	1	2	3	4	5	6
$p(X=x)$	$p/2$	$\frac{1-p}{4}$	$\frac{1-p}{4}$	$\frac{1-p}{4}$	$\frac{1-p}{4}$	$p/2$

$E(X) = \frac{7}{2}$

$E(X^2) = \frac{p}{2}(1^2 + 6^2) + \frac{1-p}{4}(2^2+3^2+4^2+5^2)$

$= 5p + \frac{54}{4}$

$\text{Var}(X) = E(X^2) - (EX)^2 = 5p + \frac{54}{4} - (\frac{7}{2})^2$

$= 5 \cdot p + \frac{5}{4}$

$SD(X) = \sqrt{\text{Var}(X)}$