Proof.

$$\begin{aligned}
var[aX+b] &= E[((aX+b)-E[aX+b])^2] \\
&= E[(aX+b-aE[X]-b)^2] \\
&= E[(aX-aE[X])^2] \\
&= E[a^2(X-E[X])^2] \\
&= a^2E[(X-E[X])^2] \\
&= a^2var[X]
\end{aligned}$$

□

Theorem. *For any random variable X*

$$var[X] = E[X^2] - (E[X])^2$$

Proof.

$$\begin{aligned}
var[X] &= E[(X-E[X])^2] \\
&= E[X^2 - 2E[X]X + (E[X])^2] \\
&= E[X^2] - 2E[X]E[X] + (E[X])^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}$$

□

example - Suppose that a box contains 10 disks with radii 1, 2, ..., and 10 respectively. One disk is selected at random.

1. Calculate the expected area of the disk
2. Calculate the expected value for the circumference and the variance for the circumference of the disk.

Let R bet the radius of the selected disk.

$$P_R(X) = \begin{cases} \frac{1}{10}, x = 1, 2, ..., 10 \\ 0, otherwise \end{cases}$$

(i) Let A be the area of the disk $A = \pi R^2$

$$
\begin{aligned}
E[A] &= E[\pi R^2] \\
&= \pi E[R^2] \\
&= \pi \left[\sum_r r^2 P_R(r) \right] \\
&= \pi \left[\sum_{r=1}^{10} r^2 \cdot \frac{1}{10} \right] \\
&= \frac{\pi}{10}(1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 + 100) \\
&= 38.5\pi
\end{aligned}
$$

(ii) Let C be the circumference $C = 2\pi R$

$$
\begin{aligned}
E[C] &= E[2\pi R] \\
&= 2\pi E[R] \\
&= 2\pi \left(\sum_{r=1}^{10} r^2 \cdot \frac{1}{10} \right) \\
&= \frac{1}{5}(\pi(55)) \\
&= 11\pi
\end{aligned}
$$

$$
\begin{aligned}
var[C] &= E[C^2] - (E[C])^2 \\
&= E[4\pi^2 R^2] - (E[C])^2 \\
&= 4\pi^2 E[R^2] - (11\pi)^2 \\
&= 4\pi^2(38.5) - 121\pi^2
\end{aligned}
$$

Theorem 3.4, Chebyshev's Inequality. *Let X be a random variable with $E[X] = \mu$ and $\sigma = \sqrt{var[X]}$.*

$$
P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, wheneverk > 0
$$

$$
\begin{aligned}
|X - \mu| < k\sigma &\Leftrightarrow -k\sigma < X - \mu < k\sigma \\
&\Leftrightarrow \mu - k\sigma < X < \mu + k\sigma
\end{aligned}
$$

note - Confidence interval

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

If k=2,

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4}$$
$$\Rightarrow P(\mu - 2\sigma < X < \mu + 2\sigma) \geq \frac{3}{4}$$

X is 75% likely to be within 2 standard deviations of its expected value.

example - Suppose that a small shipping company ships 2 metric tonnes on the average day with a standard deviation of 0.625 metric tonnes.

1. Estimate the percentage of days where the tonnage shipped is between 1.06259-2.4375.
2. Find the smallest possible interval of shipping tonnage such that the tonnage shipped has a 95% chance of being on that interval.

(i) Let X be the tonnage shipped on a particular day

$$E[X] = 2 = \mu, \sigma = 0.625$$

$$\begin{aligned} \mu - 1.5\sigma &= 2 - (1.5 \cdot 0.625) \\ &= 1.0625 \end{aligned}$$

$$\begin{aligned} \mu + 1.50 = 2(1.5)(0.625) = 2.9375 \\ P(1.0625 < X < 2.9375) \\ &= P(\mu - 1.5\sigma < X < \mu + 1.5\sigma) \\ &= k = 1.5 \text{in Chebyshev's inequality} \\ &\geq 1 - \frac{1}{1.5^2} = 1 - \frac{1}{\frac{9}{4}} = \frac{5}{9} \approx 55.6\% \end{aligned}$$

(ii) Want $1 - \frac{1}{k^2} = .95$

$$\begin{aligned} &\Leftrightarrow k^2 = 2 \\ &\Rightarrow k = 2\sqrt{5} \approx 4.47 \end{aligned}$$

$$P(\mu - 4.47\sigma < X < \mu + 4.47\sigma) \geq .95$$
$$2 - (4.47)(0.625) \approx -0.8$$
$$2 + (4.47)(0.625) \approx 4.795$$

Interval (0, 4.79) because you can't ship negative tonnage.

HOMEWORK 3.11, 3.15-3.18, 3.21