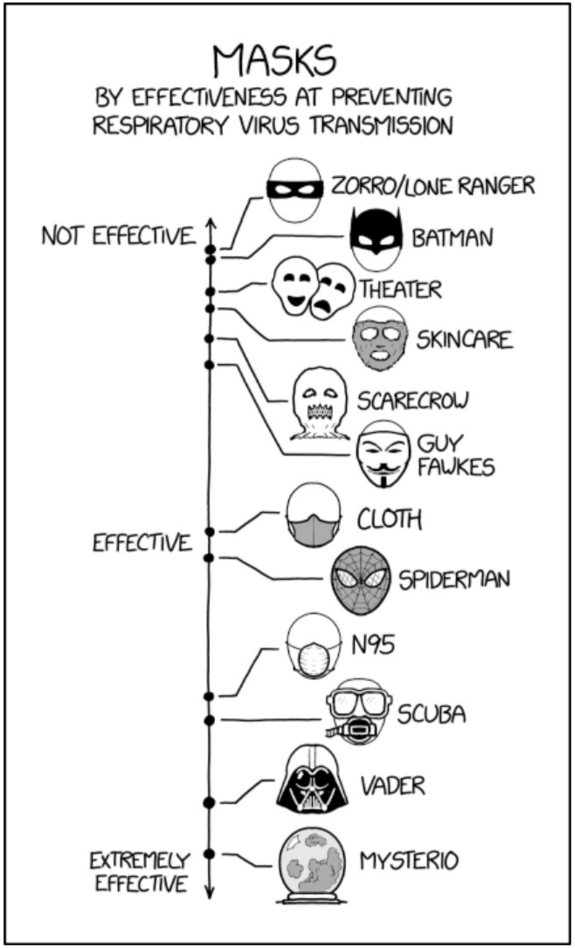
Stat 88: Probability & Math. Statistics in Data Science



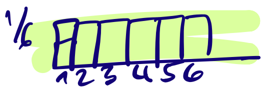
Lecture 16: Part 1

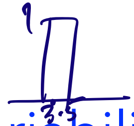
3/15/2022

Measuring Variability

Sections 6.1, 6.2

https://xkcd.com/2367/





X = # of spots when we roll a fair die
$= \{1, 2, \ldots 6\}$ w.p $1/6$
$Y = 3.5$ (Y is constant)
$\mathbb{E}(X) = \mathbb{E}(Y) = 3.5$

How should we measure variability

- The expected value (μ_X) of a random variable X is a measure of *center*.
- Expectation is a weighted average indicating the center of the distribution of mass.
- How can we describe how the values taken by the random variable ***vary*** about this center of mass? How far does a typical value land from the center?

- The difference between the values X takes and the mean is called the deviation from the mean or the average: ($D = X - \mu_X$)

$\mathbb{E}(X) = \mu_X$

- We could take each value of X, see how far it is from μ_X, and compute the (weighted) average of this distance.
- Why weighted? Values that are more likely should be counted more.

$$\mathbb{E}(D) = \mathbb{E}(X - \mu_X) = \mathbb{E}(X) - \mathbb{E}(\mu_X) = \mathbb{E}(X) - \mu_X = 0$$

Measuring variability

- Suppose that X is a rv that takes values -1 and 1 with equal probability.

$\mathbb{E}(X) = 0$

- We know what a measure of the variability should be, let's see if it works.

$D = 1, -1$ w.p $\frac{1}{2}$ each

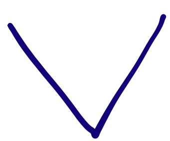
- Write out $E(D) = E((X - E(X))$. What problem do you see?
- How can we fix it?

$\mathbb{E}(D) = 0$

$\mathbb{E}(|D|)$, $|D| = \begin{cases} 1, & D = 1 \\ +1, & D = -1 \end{cases}$ w.p. $\frac{1}{2}$

$\mathbb{E}(|D|) = 1$

Note that the absolute value function $f(x) = |x|$



Example

- Consider the random variable with distribution shown below:

x	1	2	3
$P(X = x)$	0.2	0.5	0.3
$(x-\mu_X)^2$	$(1-2.1)^2 = (-1.1)^2 = 1.21$	$(2-2.1)^2 = 0.01$	$(3-2.1)^2 = 0.81$

- Find $E(X) = \mu_X$ $= 1(0.2) + 2(0.5) + 3(0.3) = 0.2 + 1 + 0.9$
  $= 2.1$
- Write down the values of $\underbrace{(x-\mu_X)^2}_{D^2}$ and find $E(X-\mu_X)^2$

$$E((X-\mu_X)^2) = (1.21)(0.2) + (0.01)(0.5) + (0.81)(0.3)$$

$$= 0.47$$

Variance of a random variable

$D = X - \mu_X$

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

$= \sum_x (x - \mu_X)^2 P(X = x)$

- The *variance* of a random variable is defined by:

$$Var(X) = E(D^2) = E[(X - E(X))^2]$$

- Note that variance is an expectation of a function of X
- We could use the absolute deviation from the mean: $|D| = |X - \mu_X|$ but it isn't as nice a function as the square of the deviation from the mean.

- The only problem with using the variance is the units are off because we squared the deviation. In order to get the proper units back, we now have to take the square root.

$SD(X) =$ Standard Deviation of X

$= \sqrt{Var(X)}$

Standard deviation of a random variable

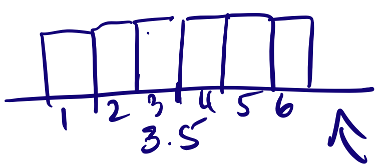
- The *standard deviation* of a random variable is the ***square root of the variance*** of the random variable.

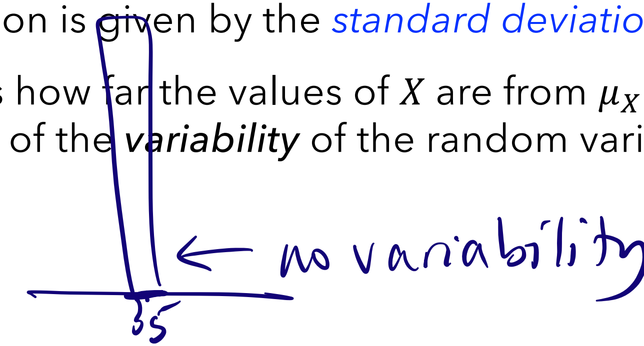
Root $\sqrt{E(D^2)}$ square

Mean

$$SD(X) = \sqrt{Var(X)} = \sqrt{E(D^2)} = \sqrt{E[(X - E(X))^2]}$$

- The variance is more convenient for computations because it doesn't have square roots. However, since the units are squared, it is difficult to interpret. Better to think about SD
- You can think of SD as the RMS deviation from the mean (***root-mean-square***)
- *Deviation* is the amount above or below the expected value. How big is it likely to be?
- The likely or typical size of the deviation is given by the *standard deviation.*
- SD is a *give-or-take* number telling us how far the values of X are from μ_X on average, that is, it gives us a measure of the ***variability*** of the random variable.





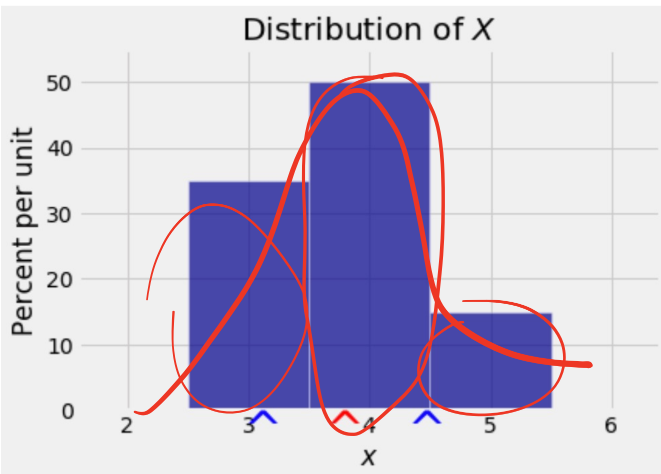
A tale (tail?) of two random variables (example from the text)

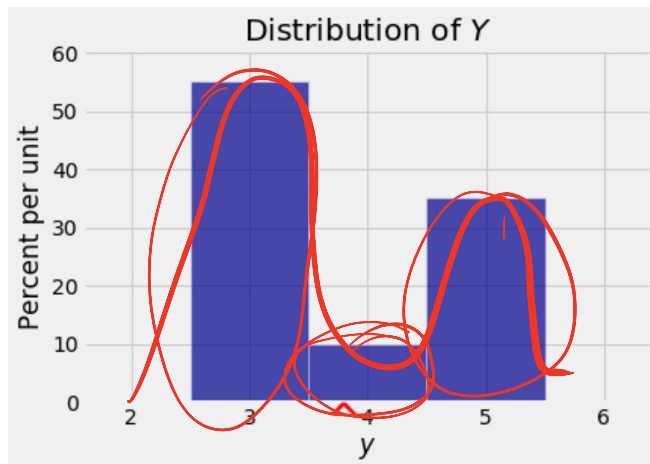
x	3	4	5
$P(X = x)$	0.35	0.5	0.15
$(x - \mu_X)^2$			

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

$E(X) = 3(0.35) + 4(0.5) + 5(0.15) = 3.8$

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





$\mathbb{E}(X) = \mathbb{E}(Y) = 3.8$

A tale (tail?) of two random variables (example from the text)

x	3	4	5
$P(X = x)$	0.35	0.5	0.15
$(x - \mu_X)^2$	$(0.8)^2$	$(0.2)^2$	$(1.2)^2$

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

- For both X and Y, compute their means and variances, then the SDs.

$$Var(X) = (0.8)^2(0.35) + (0.2)^2(0.5) + (1.2)^2(0.15)$$

$$= 0.46, \quad SD(X) \approx 0.68$$

$$Var(Y) = 0.86, \quad SD(Y) \approx 0.93$$

Comparing X and Y

- For both X and Y, compute their means and variances, then the SDs.

$$E(X) = E(Y) = 3.8$$

$$\sqrt{Var(X)} = SD(X) \approx 0.68$$

$$\sqrt{Var(Y)} = SD(Y) \approx 0.93$$

$$Var(X) = \sqrt{\sum_x (x - \mu_x)^2 \cdot P(X = x)}$$

Shortcuts and alternative formulas

- $E(X) = \sum x \cdot P(X = x) = \mu$

- $E(X^2) = \sum x^2 \cdot P(X = x)$

- $SD(X) = \sqrt{\sum (x-\mu)^2 \cdot P(X = x)}$

- $Var(X) = E[(X-\mu)^2] =$ $\sum_x (x-\mu)^2 P(X=x) = \mathbb{E}(X^2) - \mu^2$

$$= \sum_x \left[x^2 - 2\mu x + \mu^2\right] P(X=x)$$

$$= \underbrace{\sum_x x^2 \cdot P(X=x)}_{} + \sum_x (-2\mu x) \cdot P(X=x) + \sum_x \mu^2 P(X=x)$$

$$= \mathbb{E}(X^2) + \mathbb{E}(-2\mu X) + \mathbb{E}(\mu^2)$$

$$\mathbb{E}(X^2) \quad \underline{-2\mu\mathbb{E}(X)} + \underline{\mu^2}$$

$$-2\mu^2 \quad + \quad \mu^2$$

Shortcuts and alternative formulas

- If X is a random variable, and $E(X)$ is its mean
- Alternative formula for $Var(X)$:

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

$\mu = \mathbb{E}(X)$

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

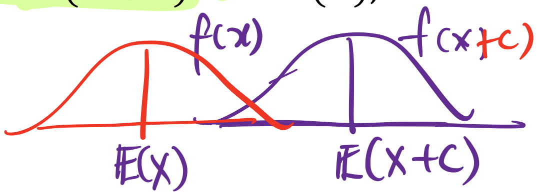
Therefore, $SD(X) = \sqrt{E(X^2) - [E(X)]^2}$

$Var(X) =$ Exp. value of $X^2 - (\mathbb{E}(X))^2$

Properties of ~~SD~~ SD

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

1. $SD(\text{constant}) = 0$ → $X = c$, $\mathbb{E}(X) = c$, Δ is always 0
   $x - \mu_x = 0$ for all x if $x = c$
   $Var(c) = 0$
2. $SD(X + c) = SD(X)$, where c is a constant



$$Var(X+c) = \mathbb{E}\left((X+c)^2\right) - \left(\mathbb{E}(X+c)\right)^2$$
$$= \mathbb{E}(X^2 + 2cX + c^2) - (\mathbb{E}(X) + 2c\mathbb{E}(X) + c^2)$$

3. $SD(cX) = |c|SD(X)$, thus $SD(-X) = SD(X)$

$$Var(cX) = \mathbb{E}((cX)^2) - (\mathbb{E}(cX))^2$$
$$= \mathbb{E}(c^2X^2) - (c\mathbb{E}(X))^2 = c^2\mathbb{E}(X^2) - c^2\mu^2$$
$$= c^2\left[\mathbb{E}(X^2) - \mu^2\right] = c^2 Var(X), \Rightarrow SD(cX) = \sqrt{c^2 Var(X)}$$
$$= |c|\, SD(X)$$

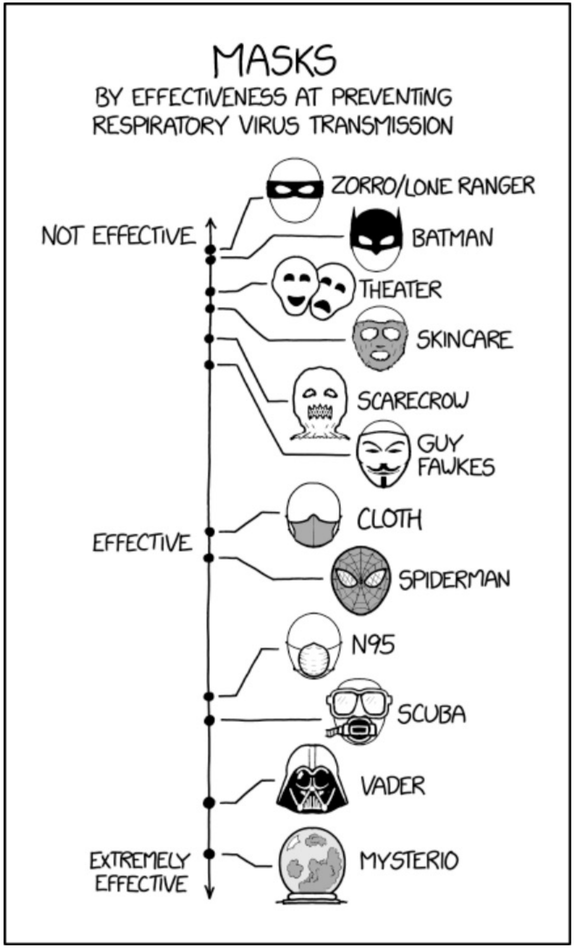
$SD(X) \geq 0$

4. If X and Y are *independent*, then $Var(X \pm Y) = Var(X) + Var(Y)$

Fact $Var(X + Y)$
$= Var(X) + Var(Y)$

$Var(X - Y) = Var(X) + Var(Y)$

Stat 88: Probability & Math. Statistics in Data Science



https://xkcd.com/2367/

Lecture 16: Part 2

3/15/2022

Markov's and Chebyshev's Inequalities

Sections 6.3, 6.4

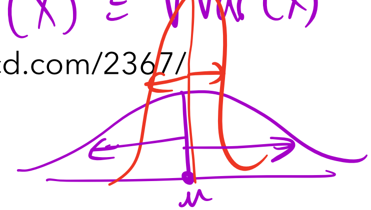
X, $f(x) = P(X=x)$, $F(x)$

$\mathbb{E}(X) = \mu$

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

$= \mathbb{E}(X^2) - \mu^2$

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



What can the average tell us?

- We want to say something about the accuracy of estimates. If we don't know much about a list beyond its mean, what can we say about the values of the random variable?

Med = Avg

1, 2, 3, 4, 5 / 1, 2, 3, 4, 5000

Med

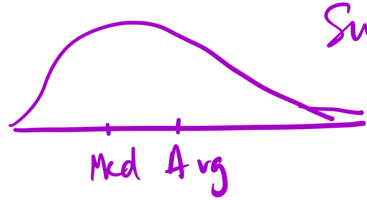
$\frac{1+2+3+4+5000}{\#5}$

- True or false? Half the data are always above the average.

No. This is true for the median

- Suppose we have 100 *non-negative* numbers and the average of the list is 2. True or false? At most 25 of the numbers could be greater than or equal to 8. Why?

$\text{Avg} = \frac{\sum x_i}{n}$



Suppose avg = 2, n = 100

Total = Avg × n = 200

Note that 8×25 = 200

- The average gives us an upper bound on proportion. That is, it puts a firm wall that prevents too many numbers from being big compared to the average.

If 25 out of 100 #s are = 8 & rest are 0, avg is 2.

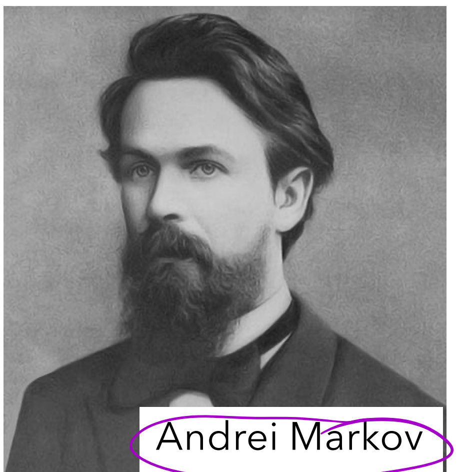
If even 26 out of 100 #s is 8, then avg > 2.

What can the average tell us?

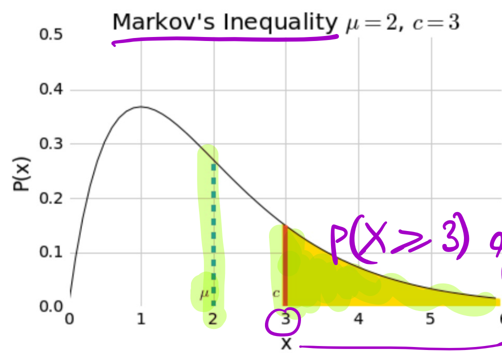
- We will begin with a simple bound, that works for non-negative random variables when the only information we have is the mean of the distribution.

Tail probabilities

- We are going to consider "tail" probabilities, or probabilities of the type $P(X \geq c)$, for some $c > 0$



Andrei Markov



$P(X \geq 3)$ given $\mu = 2$

Markov's inequality: Bounding tail probabilities

$P(X \geq c)$

- "Bounding tail probabilities" means that we put bounds on what fraction of points can fall far away from the mean.

- Let X be a non-negative random variable (all possible values taken by X are at least 0). Fix $c > 0$. We want to find an upper bound for $P(X \geq c)$in terms of $E(X)$.

$f(x) = P(X = x)$ $\quad \mu = E(X) = \sum_x x f(x)$ $\quad x \geq 0$ (given)

- Start with the definition of $E(X) = \sum_{x \geq 0} x P(X = x)$

take all x & divide into 2 groups, those $x \geq c$, those $x < c$

$a \geq \cancel{1 + 2 +} 3 + 4 + 5$

$a \geq 4 + 5$

$a \geq 2 \cdot 3 + 2 \cdot 7$

$\geq 1 \cdot 3 + 1 \cdot 7$

$E(X) = \cancel{\sum_{x<c} x \cdot f(x)} + \sum_{x \geq c} x \cdot f(x)$

$E(X) \geq \sum_{x \geq c} x \cdot f(x)$

$\geq \sum_{x \geq c} \boxed{c} f(x) = c \boxed{\sum_{x \geq c} f(x)} = c \cdot P(X \geq c)$

$\sum_{x \geq c} P(X = x)$

$$P(X \geq c) \leq \frac{E(X)}{c}$$

Markov's Ineq.: $\boxed{\frac{E(X)}{c} \geq P(X \geq c)}$

$$P(X \geq c) \leq \frac{E(X)}{c}$$

Examples

1. The mean weight of students in a certain class of students is 140 lbs. What is the largest possible fraction that could weigh over 210 lbs.?

$\mu = 140\,lb$ $\quad c = 210\ lbs$ $\quad P(X \geq c)$ how big.

$$P(X \geq 210) \leq \frac{140}{210} = \frac{2}{3} >$$ might be much larger than the true value.

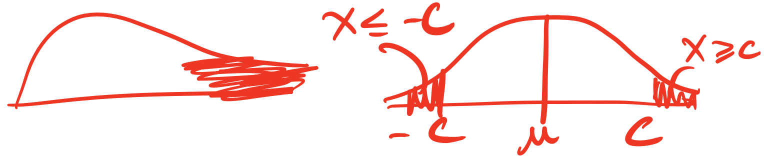
2. A student has a GPA (grade point average) of 2.8. In each course she takes, she gets a grade between 0 (failing) and 4.0 (A+). What is the largest decimal fraction of her grades that could be 4 or higher?

$E(X) = 2.8, \quad c = 4$

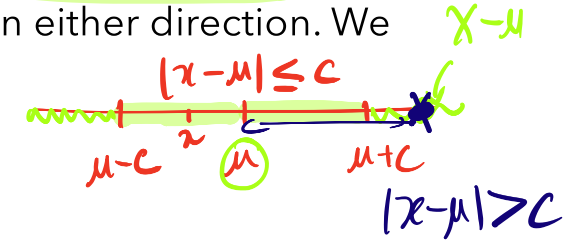
$$P(X \geq 4) = P(X = 4) \leq \frac{E(X)}{c} = \frac{2.8}{4} = 0.7$$

$\frac{2.8}{4} = 0.7$

Chebyshev's inequality



- What about if we have more information? Say we also know $SD(X)$.
- The first improvement is that we don't need X to be non-negative.
- Let $E(X) = \mu, SD(X) = \sigma$, where X is *any* random variable. Fix $c > 0$.
- We are now interested in the chance of being in either tail, so the chance of the random variable being extreme in either direction. We sum the two tail probabilities.



means dist b/w X & μ ≥ c

$$P(|X-\mu| \geq c)$$

- Now, we don't know much about $E(|X-\mu|)$ but know about the squared deviation, whose expectation is the variance of X. So we consider that random variable instead (D^2):

$P|X-\mu| \geq c) = P(\underbrace{(X-\mu)^2}_{\text{non neg. call it } Y} \geq c^2)$ What can Markov's inequality tell us about this?

X is non neg, given E(X) = μ, given c

$$P(X \geq c) \leq \frac{\mu}{c}$$

$P(Y \geq c)$, where $Y = (X-\mu)^2$, then $Y \geq 0$

$\mathbb{E}(Y) = \mathbb{E}((X-\mu)^2) = Var(X)$

Markov's ineq.

$$P(Y \geq k) \leq \frac{\mathbb{E}(Y)}{k} =$$

$$\underline{P((X-\mu)^2 \geq c^2)} \leq \frac{\mathbb{E}((X-\mu)^2)}{c^2} = \frac{Var(X)}{c^2}$$

$$P(|X-\mu| \geq c)$$

for any r.v. X, with $\mathbb{E}(X) = \mu$, $SD(X) = \sigma$

for any $c > 0$,

$$P(|X-\mu| \geq c) \leq \frac{Var(X)}{c^2} = \frac{\sigma^2}{c^2}$$

Chebyshev's inequality

- For a random variable X, with mean μ and standard deviation σ, for any positive constant $c > 0$, we have:

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} = \frac{Var(X)}{c^2}$$

- Suppose a random variable X has $\mu = 60$, and $\sigma = 5$. what is the chance that it is outside the interval $(50, 70)$?

- What about $P\left(X \in (50, 70)\right)$?

Chebyshev's inequality interpreted as distances

- Say that $E(X)$ is the origin, and we are measuring distances in terms of $SD(X)$.
- We want to know the chance that the rv X is at least $k\ SD'$s away from its mean:

- What if we are only interested in one tail? A certain type of light bulb has an average lifetime of 10,000 hours. The SD of bulb lifetimes is 550 hours. What decimal fraction of bulbs could last more than 11,980 hours?

Chebyshev or Markov?

- Suppose X is a non-negative random variable with expectation 60 and SD 5. What can we say about $P(X \geq 70)$?

Exercise 6.5.6

Ages in a population have a mean of 40 years. Let X be the age of a person picked at random from the population.

a) If possible, find $P(X \geq 80)$. If it's not possible, explain why, and find the best upper bound you can based on the information given.

b) Suppose you are told in addition that the SD of the ages is 15 years. What can you say about $P(10 < X < 70)$?

c) With the information as in Part b, what can you say about $P(10 \leq X \leq 70)$?