Stat 88 lec 21 (online class)

Warmup 2:00 - 2:10

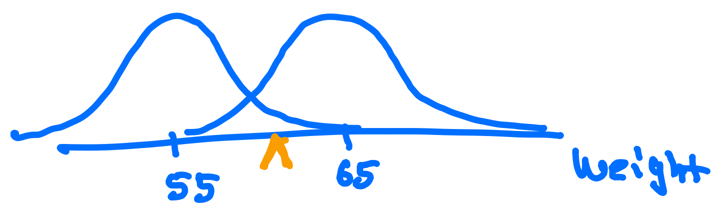
. A study on college students found that the men had an average weight of about 66 kg and an SD of about 9 kg. The women had an average weight of about 55 kg and SD of 9 kg. If you took the men and women together, would the SD of their weights be:

a) smaller than 9kg

b) just about 9 kg

c) bigger than 9kg

d) you need more information



Use "raise hand" in zoom to ask question,
You can also send me a chat message.

Last time

SD(X) is the average deviation of X from the mean, E(X).

$$SD(X) = \sqrt{E((X-\mu_X)^2)} \quad \text{where } \mu_X = E(X)$$

or

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

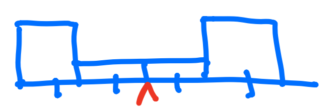
$$Var(X) = (SD(X))^2$$

You should be able to tell which of two distributions has a larger SD.

ex

4. Let X have distribution

x	1	2	3	4
$P(X=x)$	0.4	0.1	0.1	0.4

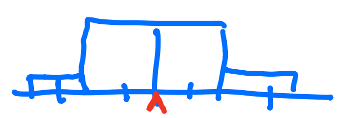


← larger SD

Which of these distributions has a larger SD?

Let Y have distribution

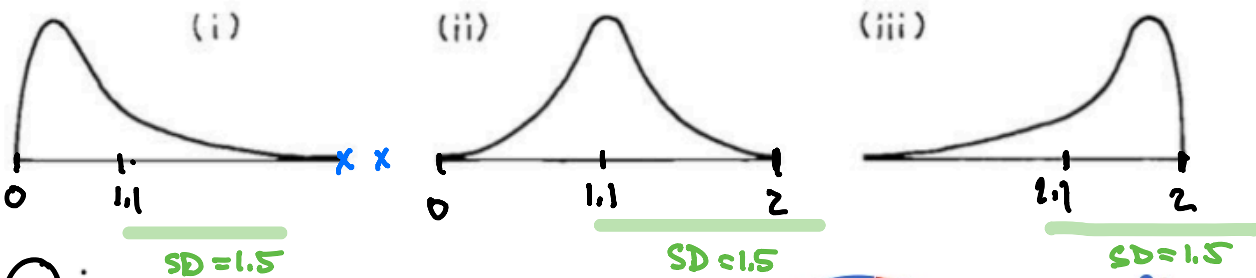
y	1	2	3	4
$P(Y=Y)$	0.1	0.4	0.4	0.1



Today ① Concept test from last time
② Sec 6.3, 6.4 Markov and Chebyshev's inequality

① Concept test from last time

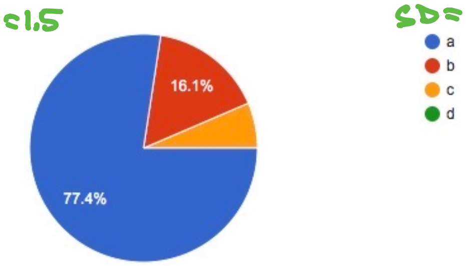
One term, about 700 Statistics 2 students at the University of California, Berkeley, were asked how many college mathematics courses they had taken, other than Statistics 2. The average number of courses was about 1.1; the SD was about 1.5. Would the histogram for the data look like (i), (ii), or (iii)? Why?



a i

b ii

c iii



b	in b, average condensed around 1.5 from mean whereas the others that are skewed left and right have smaller SDs than 15. condensed to one side
a	Because someone cannot take less than 0 math classes, the data must be skewed right (in the direction of more classes). If the data was symmetric (like the middle one), it would imply that the range is between 0 and 2.2 which is both unlikely and would mean that the SD would be much smaller than the given 1.5.

② Sec 6.3 <u>Markov's inequality</u>

What can we say about the chance a Stat 2 student takes 4 or more math classes, knowing ONLY that the average is 1.1 classes?

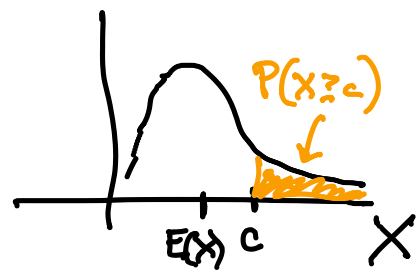
This is answered by Markov's inequality.

Tail Probabilities

Let X be a non-negative random variable.

Fix $c > 0$.

What can we say about $P(X \geq c)$ if we know $E(X)$?



$$E(X) = \sum_{\text{all } x} xP(X = x)$$

$$= \sum_{\text{all } x<c} xP(X = x) + \sum_{\text{all } x\geq c} xP(X = x)$$

nonnegative

$$E(X) \geq \sum_{\text{all } x \geq c} xP(X = x) \geq \sum_{\text{all } x \geq c} cP(X=x)$$

$$\Rightarrow \quad E(X) \geq \sum_{\text{all } x \geq c} cP(X = x)$$

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

$$= cP(X \geq c)$$

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

Markov's inequality

Ex Give an upperbound for the probability that a Stat 2 student takes 4 or more math classes. ($E(X) = 1.1$)

$$P(X \geq 4) \leq \frac{1.1}{4} = \boxed{.275}$$

ex Let X be a non-negative RV and let k be any positive constant. Find an upper bound for $P(X \geq \underbrace{kE(X)}_{c})$.

$$\leq \frac{E(\cancel{X})}{kE(\cancel{X})} = \boxed{1/k}$$

What does Markov say if $k = 1/2$?

$$P(X \geq \tfrac{1}{2}E(X)) \leq \frac{1}{\frac{1}{2}} = 2$$

ex Let $X \sim \text{Binomial}(100, 1/2)$

What is an upper bound for $P(X > 4E(X))$? $\frac{1}{4}$

What is $P(X > \underbrace{4E(X)}_{4 \cdot 50 = 200})$ exactly? $\Rightarrow$ 0 since range of X is 0 to 100.

Sec 6.4 Chebyshev's Inequality.

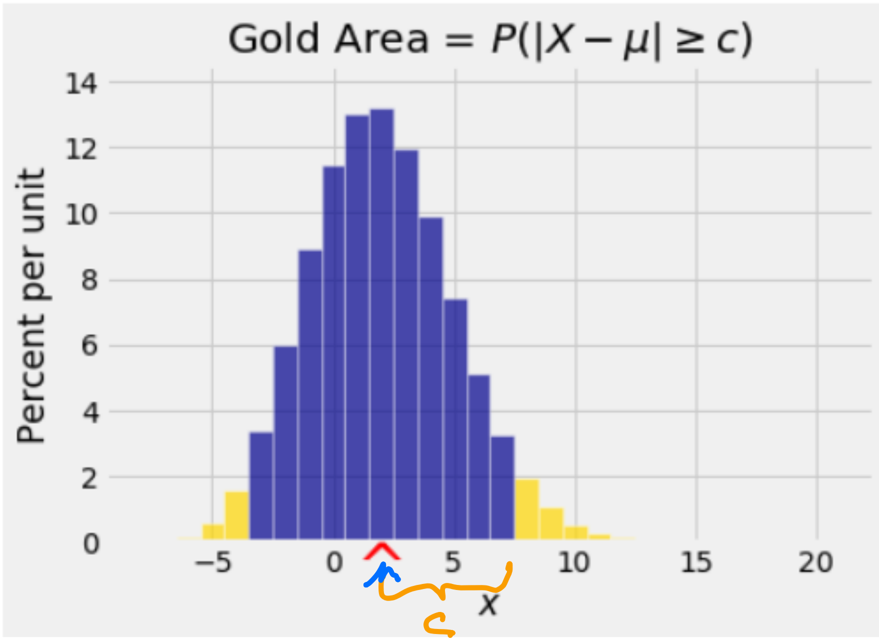
Can we get a better upper bound for the chance a Stat 2 student takes 4 or more math classes knowing the average is 1.1 classes, **AND** the SD is 1.5 classes?

This is answered by Chebyshev's inequality.

Let $\mu = E(X)$
$\sigma = SD(X)$

Let X be any RV (possibly negative) and fix $c > 0$.

We are interested in the chance of being in both tails, $P(|X-\mu| \geq c)$.



$$P(|X-\mu| \ge c) = P((X-\mu)^2 \ge c^2)$$

By Markov since RV $(X-\mu)^2 > 0$

$$\le \frac{E((X-\mu)^2)}{c^2}$$

$$= \frac{\sigma^2}{c^2}$$

Var(X)

$$\boxed{P(|X-\mu| \ge c) \le \frac{\sigma^2}{c^2}}$$

Chebyshev's inequality

ex Suppose a RV, X, has $\mu = 60$, and $\sigma = 5$.

What is the chance that it is outside the interval (50,70)?

Notice 60 is center of (50,70) $P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$

$$P(\underbrace{X \notin (50,70)}) = P(|X-60| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{5^2}{10^2} = \boxed{\frac{1}{4}}$$

What can you say about $P(\underbrace{X \in (50,70)}_{\substack{\| \\ 1 - P(X \notin (50,70)) \\ \geq \\ 1/4}})$?

$P(X \in (50,70)) \geq 3/4$ not >

Chebyshev's inequality revisited

Chebyshev's inequality can give an upper bound for the chance your data is K or more SD away from the mean. $(K > 0)$ ex $K = 2$.

Let X be a random variable with mean μ and SD σ. Then for all $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{\cancel{\sigma^2}}{k^2\cancel{\sigma^2}} = \frac{1}{k^2}$$

Chebyshev's inequality

- the chance that X is at least 2 SDs away from its mean is at most ? $1/4$
- the chance that X is at least 3 SDs away from its mean is at most ? $1/9$
- the chance that X is at least 4 SDs away from its mean is at most ? $1/16$
- the chance that X is at least 5 SDs away from its mean is at most ? $1/25$

This holds for ANY DISTRIBUTION

ex

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.

don't know distribution of X

$\leq \frac{40}{80} = \boxed{1/2}$

b) Suppose you are told in addition that the SD of the ages is 15 years. What can you say about $P(\underbrace{10 < X < 70}_{40})$?

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

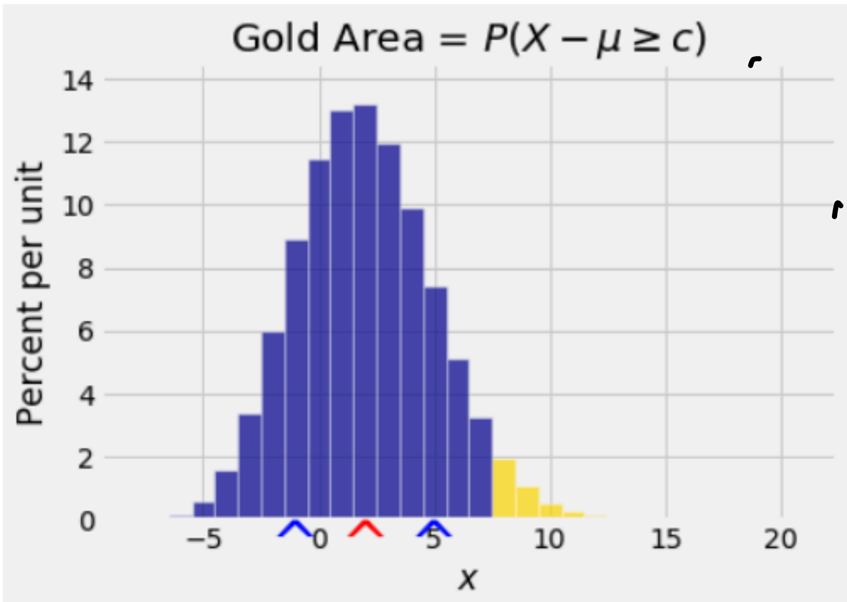
$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

$= 1 - P(|X - 40| \geq \underbrace{2 \cdot 15}_{30})$

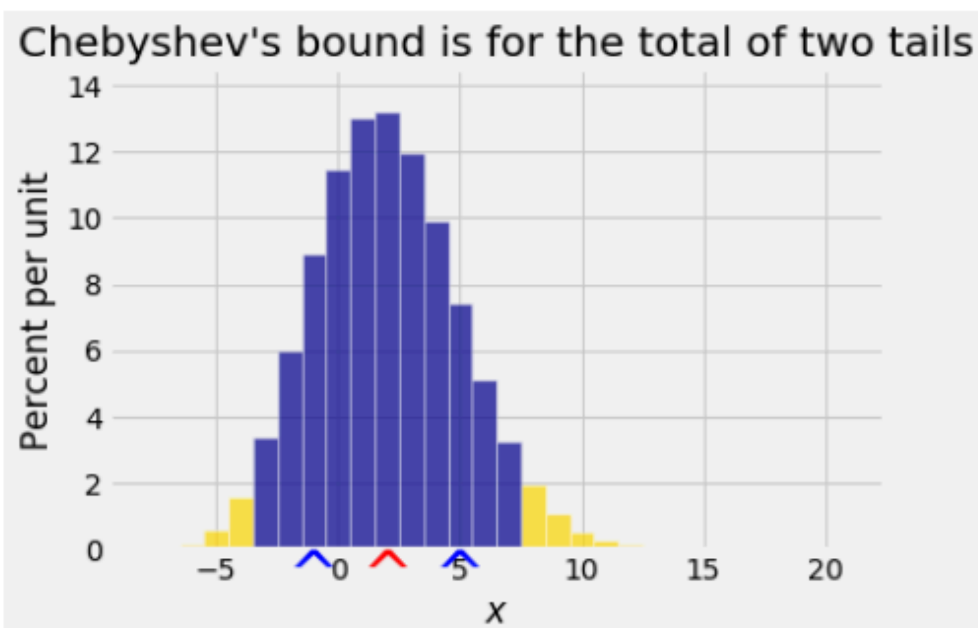
$\geq 1 - \frac{1}{2^2} = \boxed{3/4}$

Bound on one tail

Suppose we want an upper bound on just one tail, as in the figure below. The right hand tail probability is $P(X - \mu \geq c)$.



Chebyshev's inequality gives an upper bound on the total of two tails starting at equal distances on either side of the mean: $P(|X - \mu| \geq c)$.



You can't just use half of Chebyshev upper bound.

Note: each single tail is no bigger than the total of two tails.

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

ex What is chance that Stat 2 student takes 4 or more math classes given $\mu = 1.1$ and $\sigma = 1.5$

we need $X-\mu$ here to use Chebyshev

σ^2

$$P(X \geq 4) = P(X - 1.1 \geq \underbrace{4 - 1.1}_{= 2.9}) \leq \frac{(1.5)^2}{(2.9)^2} = \boxed{.267}$$

.267 .275

better upper bound.

Note that Markov gave .275 (see bottom page 5) so Chebyshev gives a better upper bound.