Stat 88 lec 30

warmup 2:00-2:10

12. A survey organization takes a simple random sample of 400 adults in a city. The annual incomes of the sampled people have an average of 68,000 dollars and an SD of 40,000 dollars.

a) Fill in the blank with one of the words "sample" or "city".

CI 's of μ ← city avg

$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$ → $68{,}000 \pm \frac{2(40000)}{20}$

$68{,}000 \pm 4000$

The interval "68,000 dollars ± 4,000 dollars" is an approximate 95% confidence interval for the average annual income of adults in the ___________. city

b) Pick all of the correct options and justify your choices. More than one option may be correct.

The normal curve used in the construction of the confidence interval in Part **a** is the distribution of

~~(i)~~ the incomes of the adults in the city

~~(ii)~~ the incomes of the adults in the sample

(iii) the averages of all possible simple random samples of 400 adults from the city

c) True or false (explain): False

The incomes of approximately 95% of the adults in the city are in the range 68,000 dollars ± 4,000 dollars.

d) Fill in the blanks with the best choices you can make from the following set. You are welcome to use the same entry more than once.

- the average income of adults in the city
- the average income of adults in the sample
- 68,000 dollars
- 40,000 dollars
- 2,000 dollars

avg income in city

$E(\bar{x}) = \mu$

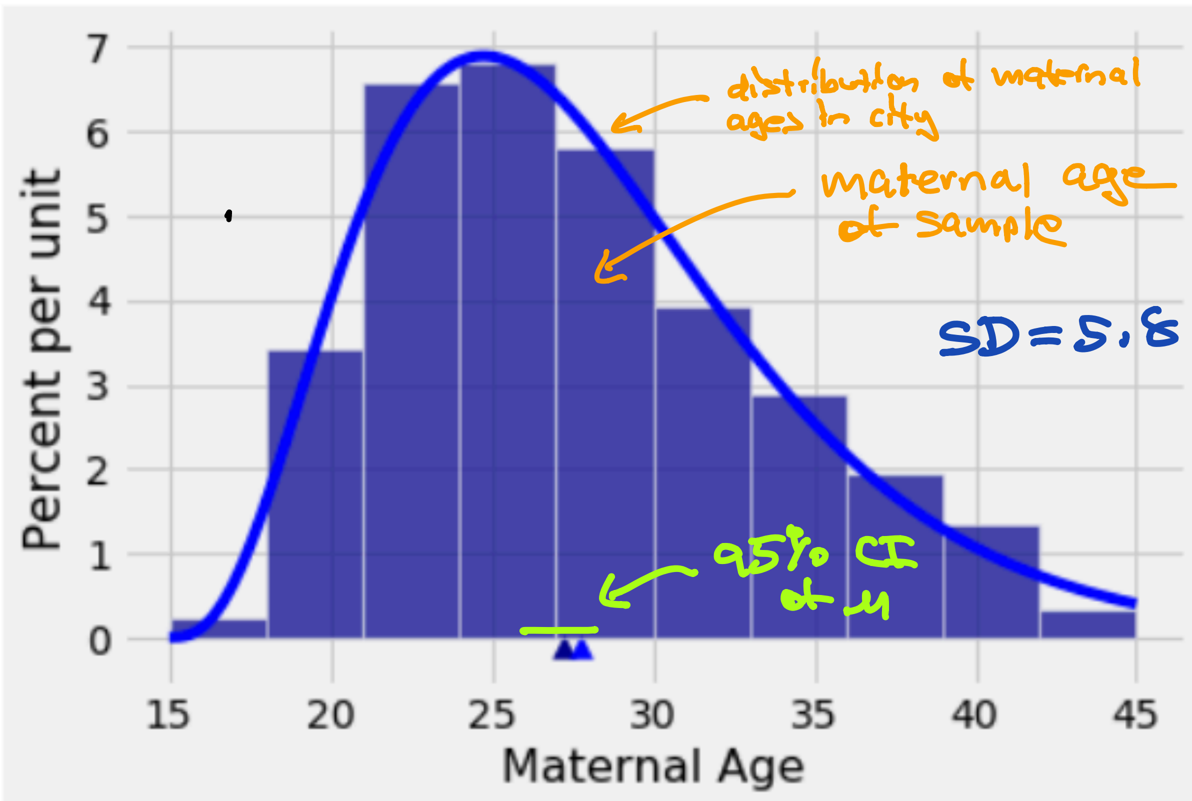
If you draw a simple random sample of 400 adults from the city, the average income of the sampled adults has expectation equal to ___________ and SD approximately equal to ___________.

$SD(\bar{x}) = \frac{\sigma}{\sqrt{n}} \approx \frac{40000}{20} = 2000$

Last time

sec 9.4 :

ex Here is a distribution of 1174 maternal ages (years) from a random sample. $\bar{X} = 27.23$, $SD(X) = 5.8$

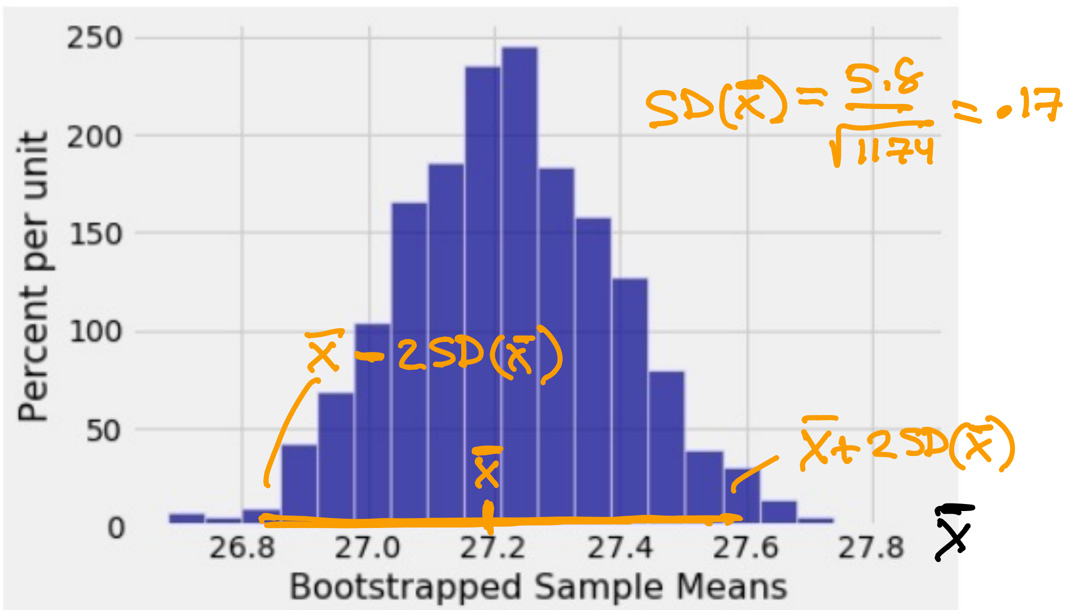


The SD of our sample of 1174 maternal ages, $SD(X) = 5.8$, is a close approximation to the population SD σ.

$SD(\bar{X}) = \frac{\sigma}{\sqrt{1174}}$ is much smaller than σ

Does this make sense? most samples have ages near the mean so we have a smaller range of ages.

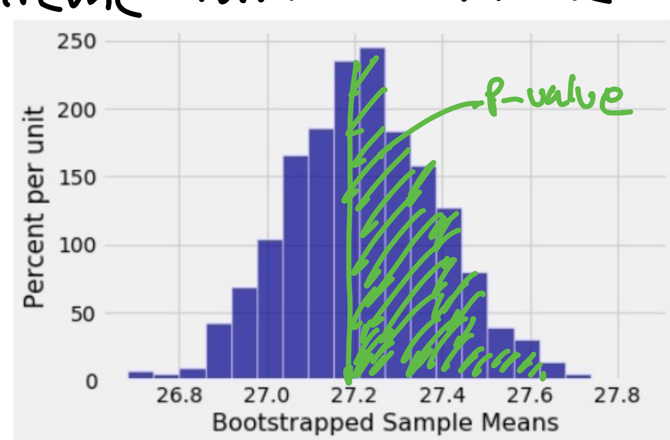
The normal curve used in the construction of your 95% CI of μ is $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}} = (26.89, 27.57)$



The p-value for the hypothesis test

$H_0: \mu = 27.4$

$H_A: \mu > 27.4$ would be the area as or more extreme than 27.4 in the direction of the alternative.



A 95% CI has a 95% chance of containing μ. Your particular 95% CI either does or does not contain μ since it is a fixed interval and μ is fixed.

Computing a 95% CI for μ is equivalent to conducting a level 5% hypothesis test:

$$\underbrace{\mu \text{ lies in } \bar{X} \pm 2\frac{\sigma}{\sqrt{n}}}_{\text{95\% CI contains } \mu}$$

iff

$$\underbrace{\bar{X} \text{ lies in } \mu \pm 2\frac{\sigma}{\sqrt{n}}}_{\text{accept } H_0 : E(X) = \mu \text{ at level 5\%}}$$

Ex Suppose $\bar{X} = 27.23$, $\sigma = 5.8$ and $n = 1174$

Test the hypothesis $H_0 : \mu = 27.4$
at level .05 $\qquad H_1 : \mu \neq 27.4$

The 95% CI for μ is $(26.89, 27.57)$ which contains 27.4 so we accept the null.

Today

① sec 10.1 density of continuous RVs,

② sec 10.2 Expectation and Variance,

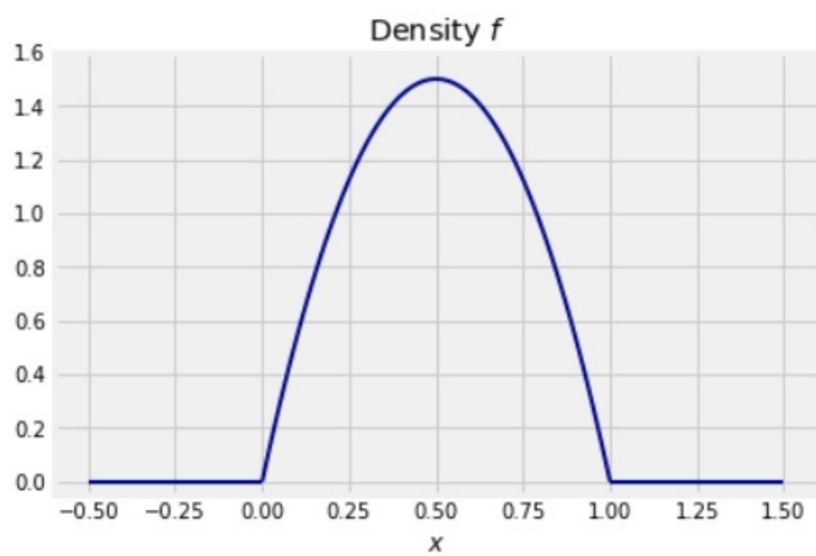
② Sec 10.1 density

A density, f, is a non negative function

s.t. $\int_{-\infty}^{\infty} f(x)dx = 1$

ex

$$f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 6x(1-x) & \text{if } 0 < x < 1 \\ 0 & \text{if } x \ge 1 \end{cases}$$

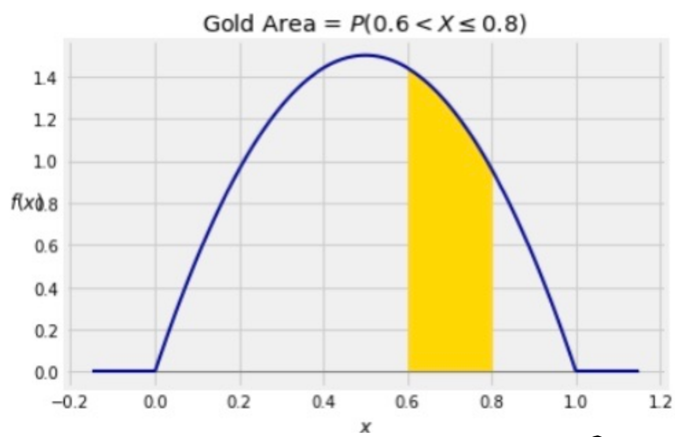


Density is not the same as probability

$f(x)$ can be > 1 (ex $f(.5) = 1.5$ above)
so $f(x)$ isn't a probability.

Areas are probabilities

A RV, X, is said to have density, f, if for every pair $a<b$ $\quad P(a < X \leq b) = \int_a^b f(x)\,dx$



Here $P(.6 < X \leq .8) = \int_{.6}^{.8} 6x(1-x)\,dx$

No probability at a single point

$P(X=b) = \int_b^b f(x)\,dx = 0 \quad$ so

$$P(a < X \leq b) = P(a < X < b)$$

We can ignore endpoints.

1. Let $X_1, X_2, X_3, \ldots$ be i.i.d. with density given by

$$f(x) = \begin{cases} 0 & x \leq 50 \\ \frac{c}{x^4} & x > 50 \end{cases}$$

This is one of the *Pareto* densities, sometimes used in economics to represent distributions of wealth in populations where a small percent of the population owns a large percent of the wealth.

a) Find c.

$$\int_{50}^{\infty} \frac{c}{x^4}\,dx = 1 \Rightarrow c = \frac{1}{\int_{50}^{\infty} \frac{1}{x^4}\,dx}$$

$$\int_{50}^{\infty} x^{-4}\,dx = \frac{1}{-3} x^{-3} \Big|_{50}^{\infty} = 0 + \frac{1}{3} \cdot \frac{1}{50^3}$$

times

$$\Rightarrow \boxed{c = 3 \cdot 50^3}$$

Cummulative distribution function (CDF)

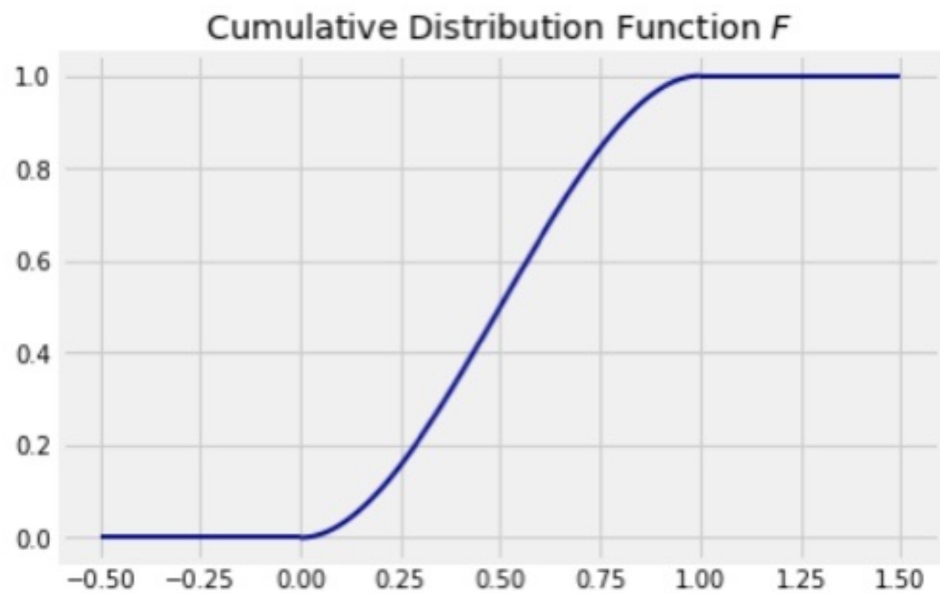
The CDF of X is $F(x) = P(X \leq x)$

$$= \int_{-\infty}^{x} f(s)\,ds$$

ex for $f(x) = 6x(1-x)$ for $0 < x < 1$

$$F(x) = \int_0^x 6s(1-s)\,ds = \int_0^x 6s\,ds - \int_0^x 6s^2\,ds$$

$$= 3x^2 - 2x^3$$



As with the normal CDF

$$P(a < X < b) = F(b) - F(a)$$

ex $P(.6 < X < .8) = \int_{.6}^{.8} \underbrace{6x(1-x)}_{f(x)}dx = \underbrace{3x^2 - 2x^3}_{= F(x)} \Big|_{.6}^{.8}$

$= F(.8) - F(.6)$

recall by FTC $\frac{d}{dx} F(x) = \frac{d}{dx} \int_{-\infty}^{x} f(s)ds = f(x)$

exercise 10.5.1

. Let $X_1, X_2, X_3, \ldots$ be i.i.d. with density given by

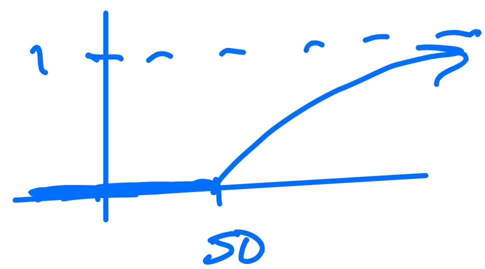
$$f(x) = \begin{cases} 0 & x \le 50 \\ \frac{c}{x^4} & x > 50 \end{cases}$$

b) Find the cdf of X_1 and sketch its graph.

$c = 3 \cdot 50^3$

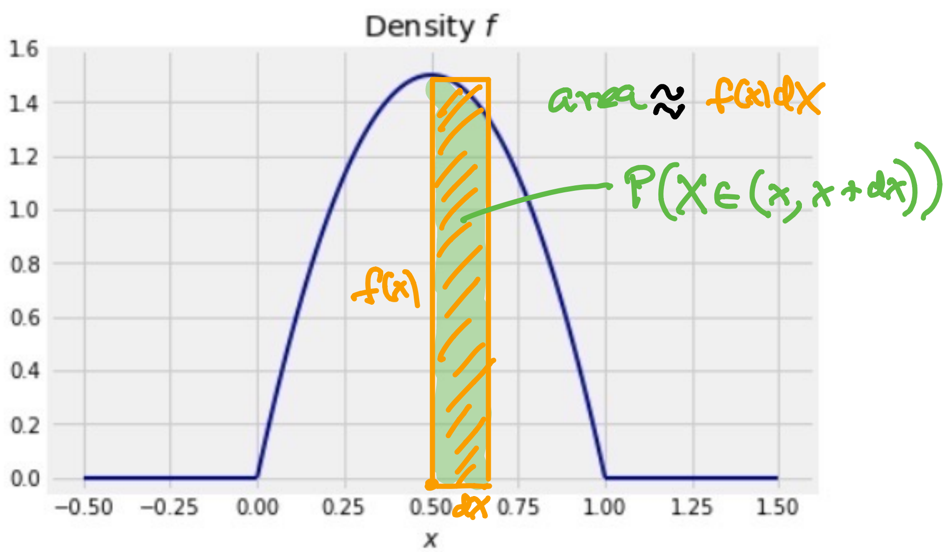
$$F(x) = 3\cdot 50^3 \int_{50}^{x} \frac{1}{s^4}\,ds = 3\cdot 50^3 \left.\frac{1}{-3s^3}\right|_{50}^{x}$$

$$= 3\cdot 50^3 \left[\frac{1}{3x^3} + \frac{1}{3\cdot 50^3}\right] = \boxed{1 - \left(\frac{50}{x}\right)^3}$$



The meaning of density

$f(x)\,dx$ is a probability



$$f(x) \approx \frac{P(X \in (x, x+dx))}{dx}$$ ← % in bin ← width of bin

i.e. f measures the chance that X is in a tiny interval near x, relative to the width of the interval.

② Sec 10.2 Expectation and Variance

If RV X has density f,

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

discrete case

$$E(X) = \sum_{x=-\infty}^{\infty} x\, P(X=x)$$

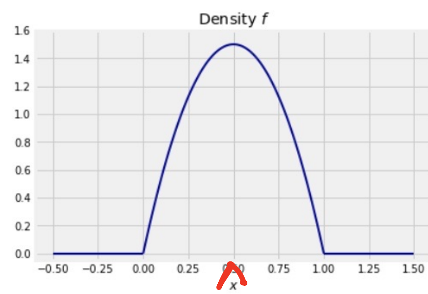
$$E(X^2) = \sum_{x=-\infty}^{\infty} x^2 P(X=x)$$

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

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

ex

$$f(x) = \begin{cases} 6x(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$



From graph of $f(x)$

$E(X) = .5$

Find $\mathrm{Var}(X)$

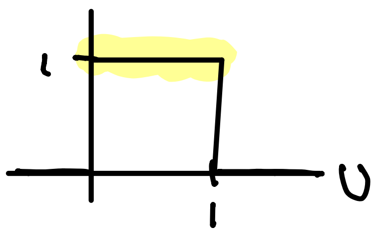
$$\int_{-\infty}^{0} x^2 \cdot 0\,dx =$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2 6x(1-x)\,dx = 6\int_0^1 (x^3 - x^4)\,dx$$

$$= 6 \cdot \tfrac{1}{4} - 6 \cdot \tfrac{1}{5} = \boxed{.3}$$

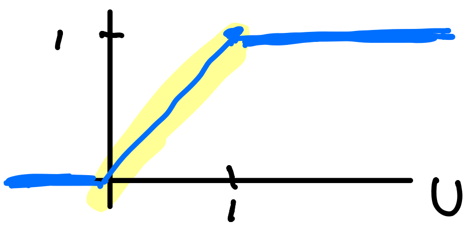
$$\mathrm{Var}(X) = .3 - (.5)^2 = \boxed{.05}$$

Uniform (0,1) (std uniform)



$$f(u) = \begin{cases} 1 & 0<u<1 \\ 0 & \text{else} \end{cases}$$

sketch the cdf of f, $F(u)$



what is the antiderivative of $f(u)=1$?

answ $F(u)=u$

$E(U)=\frac{1}{2}$

Find Var (U)

$$E(U^2) = \int_0^1 u^2 \cdot 1\, du = \frac{u^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$\text{Var}(U) = E(U^2) - (E(U))^2$$

$$= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \boxed{\frac{1}{12}}$$