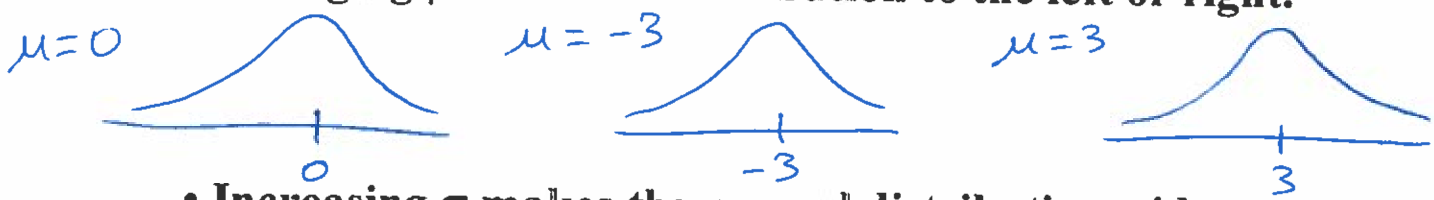
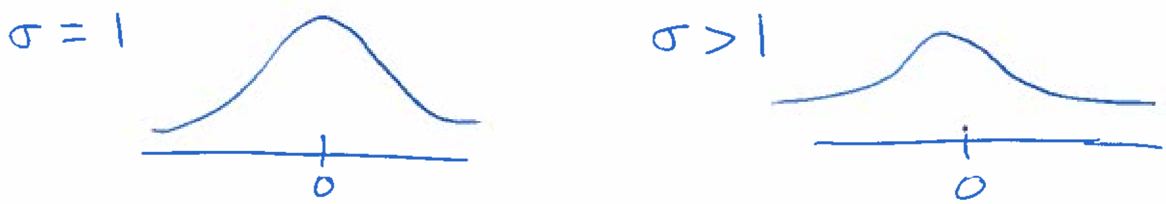
Finding Probabilities for *any* Normal r.v.

Note: There are many different normal distributions (change μ and/or σ, get a different distribution).

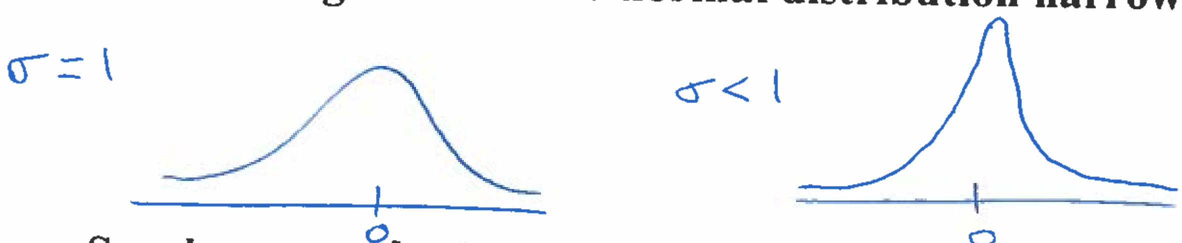
- **Changing μ shifts the distribution to the left or right.**



- **Increasing σ makes the normal distribution wider.**



- **Decreasing σ makes the normal distribution narrower.**



So why so much emphasis on the <u>standard</u> normal?

<u>Standardizing</u>: If a r.v. X has a normal distribution with mean μ and standard deviation σ, then the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.

So: We can convert <u>any</u> normal r.v. to a standard normal and then use Table II to find probabilities!

Example: Assume lengths of pregnancies are normally distributed with mean 266 days and standard deviation 16 days. $\mu = 266, \quad \sigma = 16$

What proportion of pregnancies last less than 255 days?

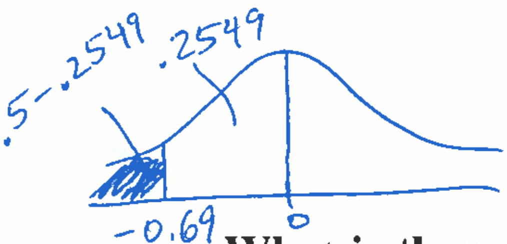
X = pregnancy length $\qquad P(X < 255)$

Standardize:

$$x = 255 \Rightarrow z = \frac{255 - 266}{16} = -0.69$$

$$P(X < 255) = P(Z < -0.69) = .5 - .2549$$

$$= \boxed{.2451}$$



What is the probability that a random pregnancy will last between 260 and 280 days? $P(260 < X < 280)$

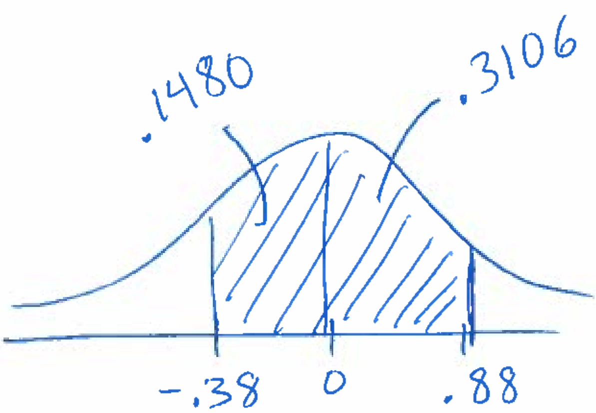
$$x = 260 \Rightarrow z = \frac{260 - 266}{16} = -0.38$$

$$x = 280 \Rightarrow z = \frac{280 - 266}{16} = 0.88$$

$$P(260 < X < 280) = P(-0.38 < Z < 0.88)$$

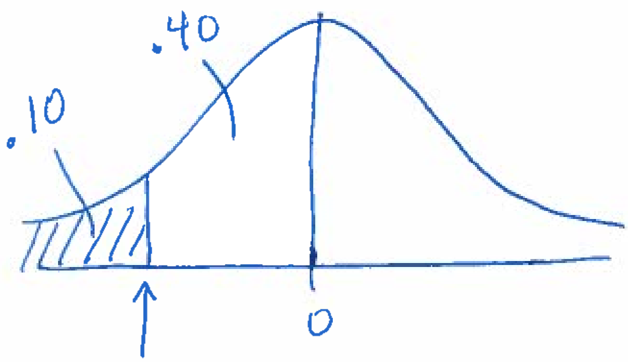
$$= .1480 + .3106$$

$$= \boxed{.4586}$$



We can also find the particular value of a normal r.v. that corresponds to a given proportion.

Example: Suppose the shortest tenth of pregnancies are classified as "unusually premature." What's the maximum pregnancy length that would be classified as such?



$z = -1.28$

Closest area inside the table to 0.40 is 0.3997 → corresponds to z-value of −1.28

We need to "unstandardize" to get back to the *X* value (pregnancy length).

$$Z = \frac{X - \mu}{\sigma} \Rightarrow -1.28 = \frac{X - 266}{16}$$

$$\Rightarrow (-1.28)(16) = X - 266$$

$$X = 266 - (1.28)(16)$$

$$X = 245.52 \text{ days}$$

General Rule: To unstandardize a z-value, use:

$$X = Z\sigma + \mu$$

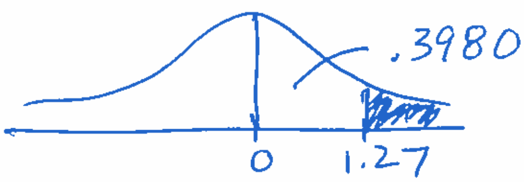
More Normal Probabilities

Example: Suppose newborn babies' weights are normally distributed with mean 7.6 pounds and std. deviation 1.1 pounds. $X = \text{weight (in pounds)}$

- **What proportion of babies are greater than 9 pounds at birth?** $P(X > 9) = ?$

$$x = 9 \Rightarrow z = \frac{9 - 7.6}{1.1} = 1.27$$

$$P(X > 9) = P(Z > 1.27) = .5 - .3980$$



$$= \boxed{.1020}$$

- **What is the probability that a randomly selected newborn is between 5 and 6 pounds?**

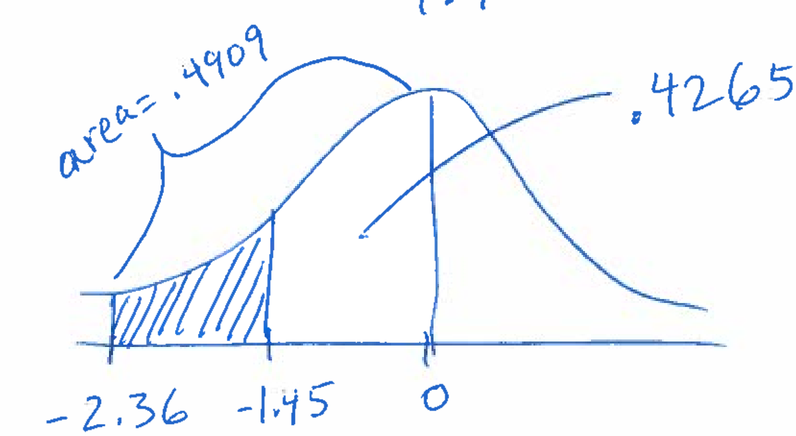
$$P(5 < X < 6) = P(-2.36 < Z < -1.45)$$

$$x = 5 \Rightarrow z = \frac{5 - 7.6}{1.1} = -2.36$$

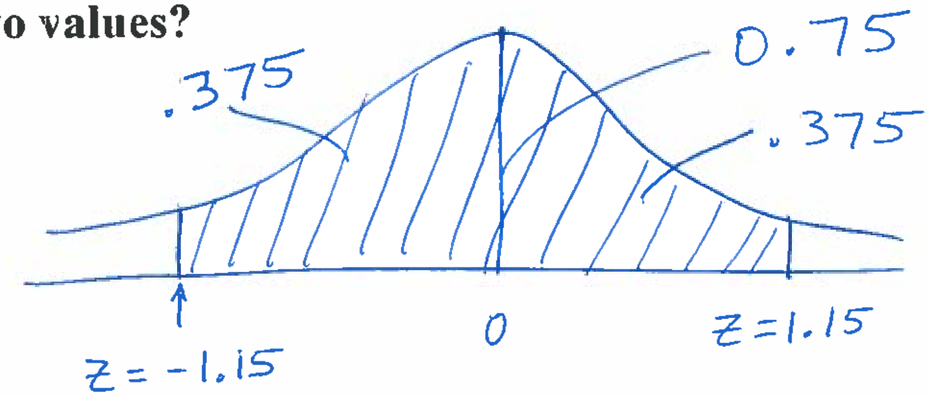
$$x = 6 \Rightarrow z = \frac{6 - 7.6}{1.1} = -1.45$$

$$= .4909 - .4265$$

$$= \boxed{.0644}$$



• **The middle 75% of babies' weights are between what two values?**



$$X = Z\sigma + \mu$$

$z = -1.15 \Rightarrow \quad X = (-1.15)(1.1) + 7.6$

$X = 6.335$ pounds

$z = 1.15 \Rightarrow \quad X = (1.15)(1.1) + 7.6$

$X = 8.865$ pounds

• **The normal model is not appropriate for every data set.**
• **It tends to give a decent approximation to the behavior of many variables observed in nature.**
• **Why? Many natural phenomena are in fact the sum total of lots of different factors that act independently to produce the final value.**
• **We will see that the normal distribution can be theoretically justified as a model for the sum of many independent quantities.**

Normal Approximation to the Binomial

The normal distribution is very powerful --- can be used to approximate probabilities for r.v.'s that are not normal.

Calculating binomial probabilities using Table I: doesn't cover all values of *n* (only 5-10, 15, 20, 25)

Using the binomial probability formula can be tedious for large *n*.

Fortunately, when *n* is large, the binomial distribution closely resembles the normal distribution with mean *np* and standard deviation $\sqrt{npq}$.

<u>Rule of Thumb</u>: When can this normal approximation be applied?

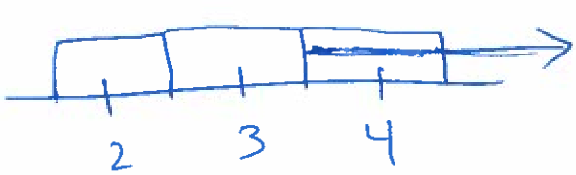
When $np - 3\sqrt{npq} \geq 0$

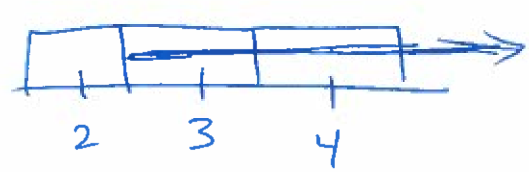
and

$np + 3\sqrt{npq} \leq n$

<u>Continuity Correction</u>: Since the normal is a continuous distribution and the binomial distribution a discrete distribution, an adjustment of 0.5 is usually made to the value of interest.

$P(X>3)$ ~~3.5~~ $\qquad$ $P(X \geq 3)$





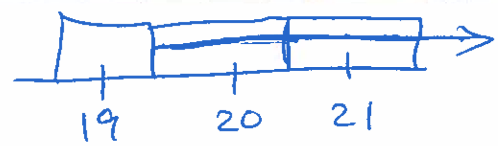
$3 \longrightarrow 3.5$ $\qquad$ $3 \longrightarrow 2.5$

Example: A hotel has found that 5 percent of its guests will steal towels. If there are 220 rooms with guests in a hotel on a certain night, what is the probability that at least 20 of the rooms will need the towels replaced?

$n = 220, \; p = .05, \; q = .95$ $\quad$ X = # rooms needing towels replaced

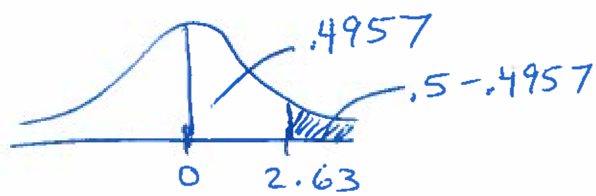
Check: $np = 220(.05) = 11$ $\quad$ $\sqrt{npq} = \sqrt{(220)(.05)(.95)} = 3.23$ $\Rightarrow$ $11 - 3(3.23) = 1.31 \geq 0$ ✓ $\quad$ $11 + 3(3.23) = 20.69 \leq 220$ ✓

$P(X \geq 20) \approx P(X^* > 19.5) \rightarrow P(Z > 2.63) \approx \boxed{.0043}$



where X^* is normal with mean 11 and std. dev. 3.23.

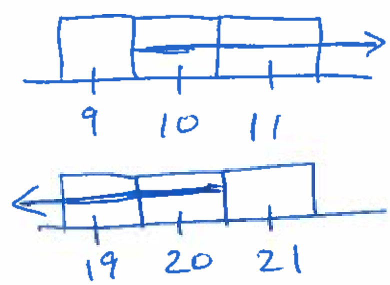
$$X^* = 19.5 \Rightarrow Z = \frac{19.5 - 11}{3.23} = 2.63$$



What is the probability that between 10 and 20 rooms will need towels replaced?

$P(10 \leq X \leq 20)$

$\approx P(9.5 < X^* < 20.5) \rightarrow P(-0.46 < Z < 2.94)$



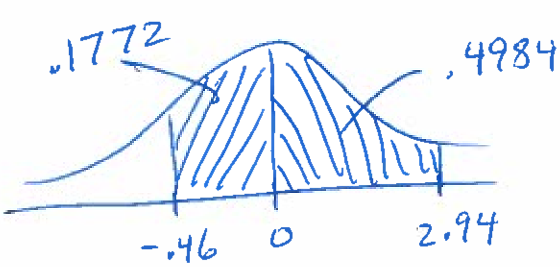
X^* is normal with mean 11 and std. dev. 3.23.

$$X^* = 9.5 \Rightarrow Z = \frac{9.5 - 11}{3.23} = -0.46$$

$$X^* = 20.5 \Rightarrow Z = \frac{20.5 - 11}{3.23} = 2.94$$

$= .1772 + .4984$

$= \boxed{.6756}$



The Exponential Distribution

Often, a waiting time (a continuous and positive r.v.) can be modeled with an exponential distribution:

pdf: $f(x) = \frac{1}{\theta} e^{-x/\theta}$ for $x > 0$ and $\theta > 0$

- Here, θ is the mean of the exponential distribution.

For example, suppose an exponential r.v. X is the waiting time (in days) between accidents at a plant. Then θ is the expected waiting time between accidents.

The mean waiting time is
$\mu = E(X) = \theta$

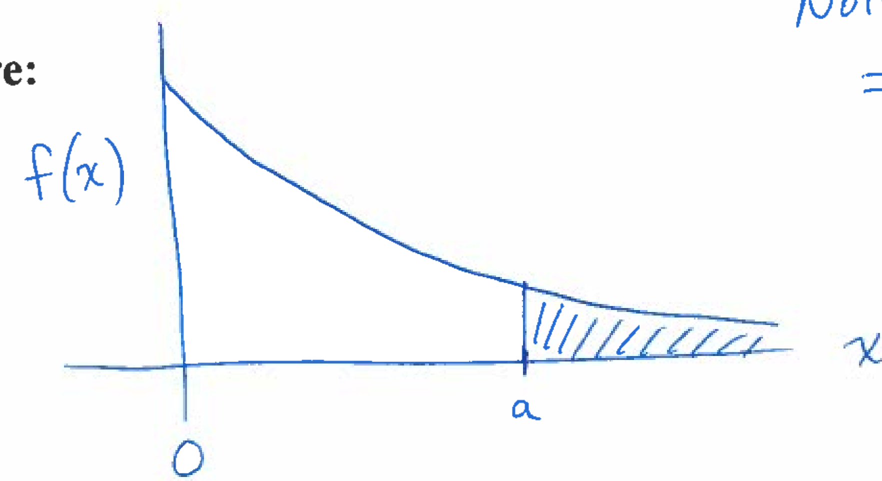
The variance is θ^2 $\rightarrow$ Standard deviation = θ

Example: Suppose that the waiting time between accidents at a plant follows an exponential distribution with mean 20.

In general: $P(X > a) = e^{-a/\theta}$

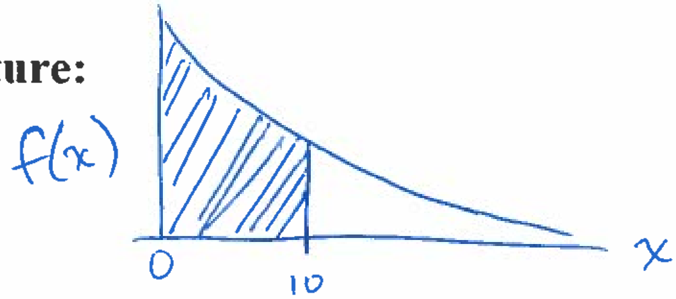
Note: $P(X \leq a)$
$= 1 - e^{-a/\theta}$

Picture:



What is the probability that the *time between* the next two accidents will be less than 10? Note: $\theta = 20$, here.

Picture:



Solution: **P(*X* < 10) =** $1 - P(X \geq 10)$

$= 1 - e^{-10/20}$

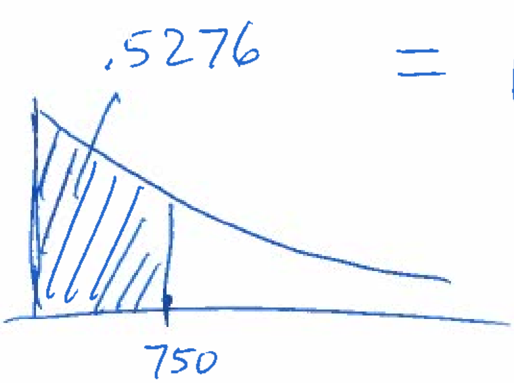
$= 1 - e^{-0.5} = \boxed{.3935}$

Example: If the time to failure for an electrical component follows an exponential distribution with a mean failure time of 1000 hours. What is the probability that a randomly chosen component will fail before 750 hours? $X \sim \text{expon}(\theta = 1000)$

$P(X < 750) = 1 - P(X \geq 750)$

$= 1 - e^{-750/1000}$

$= 1 - e^{-0.75} = \boxed{.5276}$



Relationship between Exponential and Poisson r.v.'s

- **Suppose the Poisson distribution is used to model the probability of a specific number of events occurring in a particular interval of time or space.**
- **Then the time or space between events is an *exponential* random variable.**

Why? Suppose we have a Poisson distribution with mean λt, where t is a particular length of time.

What is the probability that no events will occur before time t?

$$P[X=0] = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

Note: This situation implies that the waiting time T before the next event is more than t.

Therefore $P[T > t]$ = $e^{-\lambda t} = e^{-t/\theta}$ where $\lambda = \frac{1}{\theta}$

And so T must follow an exponential distribution with mean $\theta = 1 / \lambda$.

Example: Suppose the number of machine failures in a given interval of time follows a Poisson distribution with an average of 1 failure per 1000 hours. $\Rightarrow \lambda = \frac{1}{1000}$

waiting time between failures

T is expon$(\theta = 1000)$

What is the probability that there will be no failures during the next 2000 hours?

$$P(T > 2000) = e^{-2000/1000}$$

$$= e^{-2} = \boxed{.1353}$$

What is the probability that the time until the next failure is between 2000 and 2200 hours?

$$P[2000 \le T \le 2200]$$

$$= P[T \ge 2000] - P[T \ge 2200]$$

$$= e^{-2000/1000} - e^{-2200/1000} = e^{-2} - e^{-2.2} = \boxed{.0245}$$

What is the probability that more than 2500 hours will pass before the next failure occurs?

$$P[T > 2500]$$

$$= e^{-2500/1000} = e^{-2.5} = \boxed{.0821}$$