Homework: Chapter 8, # 45b, 59b, 67a

Sections 8.5 to 8.7: CONTINUOUS RANDOM VARIABLES

Find probabilities for *intervals*, not single values.

X = a continuous random variable, can take any value in one or more intervals.

$P(a < X < b)$ = proportion of the population with values in interval (a to b).

We consider 3 cases:

1. Uniform random variable
2. Normal random variable
3. Normal approximation for a binomial random variable

For each of these, you should be able to find probabilities like the following, where *a* and *b* are given numbers and X is a random variable of specified type:

$P(a < X < b)$

$P(X < a)$

$P(X > b)$

Note that for continuous random variables, $>$ and $\geq$ are the same because the probability of X equally an exact value is essentially 0.

For a discrete random variable (such as binomial) approximated by normal, that's not the case. It will matter whether it is $>$ or $\geq$.

UNIFORM RANDOM VARIABLES

EX: What time of day a child is born.

Let X = exact time a randomly selected child is born (Natural, not Cesarean!). Assume equally likely to be anytime in 24 hours.

SO, e.g. P(midnight < X < 6am) = 6 hours/24 hours = ¼.

We can draw a picture to illustrate this. (Picture and formulas on board.)

GENERAL DEFINITION OF CONTINUOUS RANDOM VARIABLE:

The probability density function (A different pdf abbreviation!) for a *continuous* random variable X is denoted as f(x), and is the formula for a curve such that:

1. Total area under the curve = 1
2. P(a < X < b) = area under the curve between a and b.

NOTATION : f(x) is the pdf for the random variable X. It is a function such that

$$P(a < X < b) = \int_a^b f(x)dx$$

The mean μ, variance σ^2 and standard deviation σ for X are given by:

$$\mu = \int_{-\infty}^{\infty} xf(x)dx \qquad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \qquad \sigma = \sqrt{\sigma^2}$$

Won't need calculus, will use tables and R commander (and/or Excel).

UNIFORM DISTRIBUTION between *c* and *d*:

f(x) = $\dfrac{1}{d-c}$ and area between any two numbers *a* and *b* is $\dfrac{b-a}{d-c}$

Ex: X = birth time, *c* = 0 and *d* = 24, f(x) = 1/24; P(midnight to 6am) = P(0<X<6) = 6/24 = 1/4.

Mean and standard deviation for a uniform random variable on board

NORMAL RANDOM VARIABLES

Completely defined once you know mean and standard deviation. f(x) on board

Write down the following for yourself:
1. How many hours of sleep you got last night.
2. Your height.
3. Your verbal SAT score.

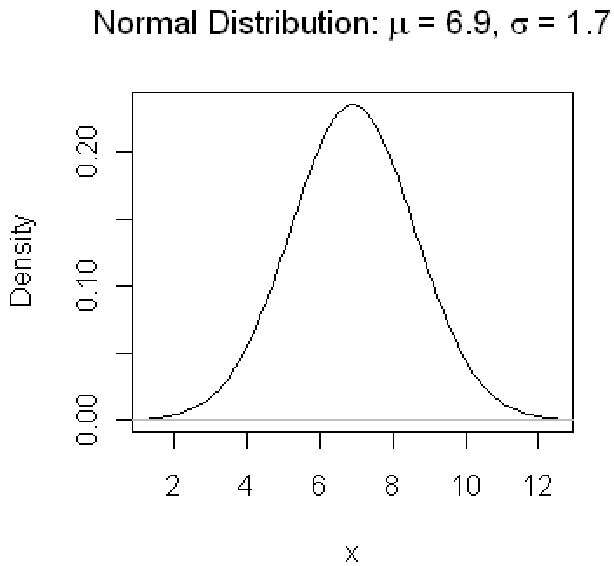
These are all approximately normal random variables, so you can determine where you fall relative to everyone else if you know μ and σ.

Random variable:	μ	σ
Sleep hours for students:	6.9 hours,	1.7 hours
Women's heights:	65 inches	2.7 inches
Men's heights	70 inches	3.0 inches
Verbal SAT scores for UCI students	569*	75
Verbal SAT scores for population	500	100

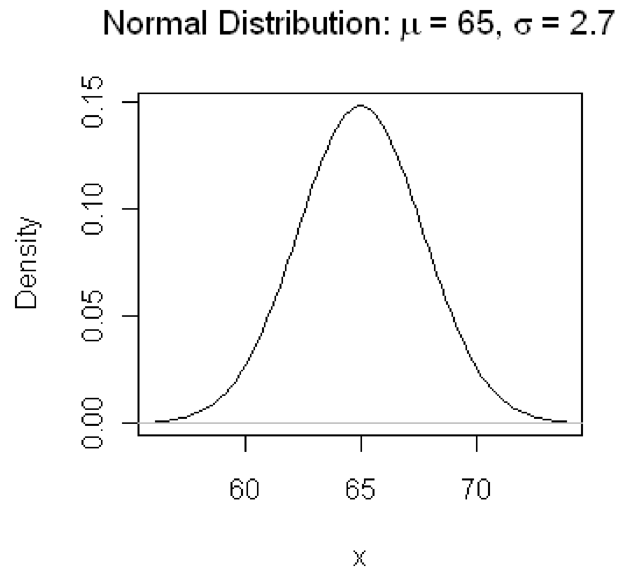
*Note that SAT means differ by school at UCI. You can see them here: http://www.oir.uci.edu/adm/IA24-fall-fr-mean-sat-by-school-2000-2008.pdf

Pictures of these:

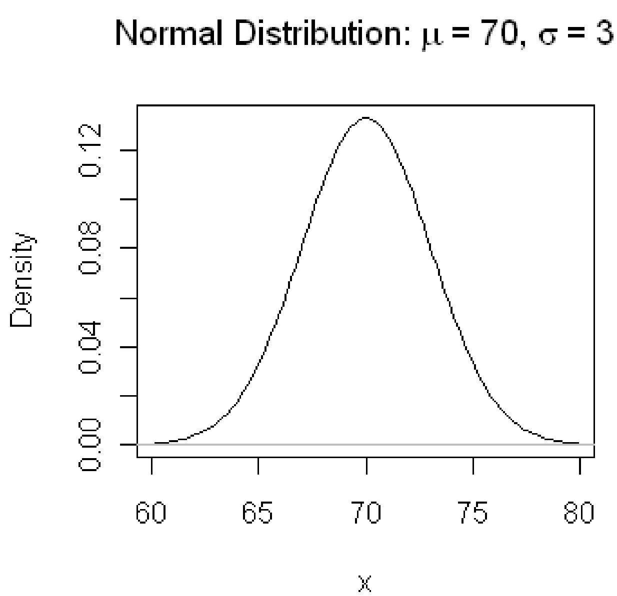
Hours of sleep



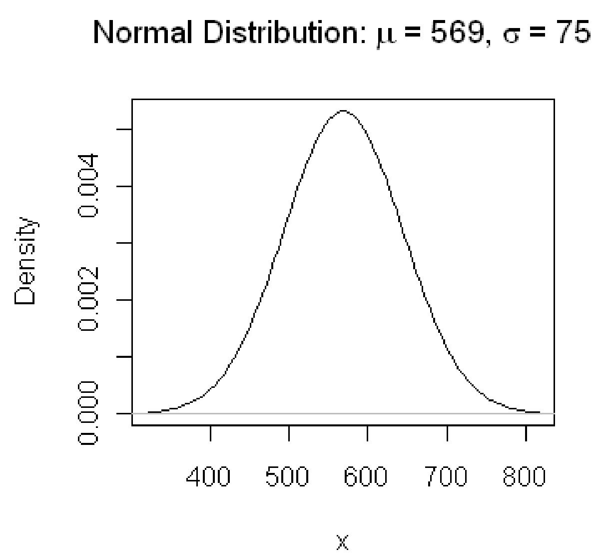
Female heights



Male heights



UCI Verbal SAT scores



What is the same and what is different about these pictures?

HOW TO FIND PROBABILITIES FOR NORMAL RANDOM VARIABLES

Two methods; in both cases you need to know mean, standard deviation, and value(s) of interest:

1. Convert value(s) of interest to z-scores, then use computer *or* Table A.1
2. Use computer directly.

Often you will need to use Rules 1 and/or 2 from Chapter 7 as well.

Always draw a picture so you know if your answer makes sense!

Method 1 (k is a value of interest; μ and σ are the mean and standard deviation):
Step 1: Convert k to a z-score, which is ***standard normal*** with $\mu=0$ and $\sigma=1$:

$$z = \frac{k - \mu}{\sigma}$$

Step 2: Look up z in Table A.1, or use R Commander or Excel.

Table A.1 gives areas *below* z. Here is a small part of the table:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

Examples (on board):

$P(z < -2.21)$

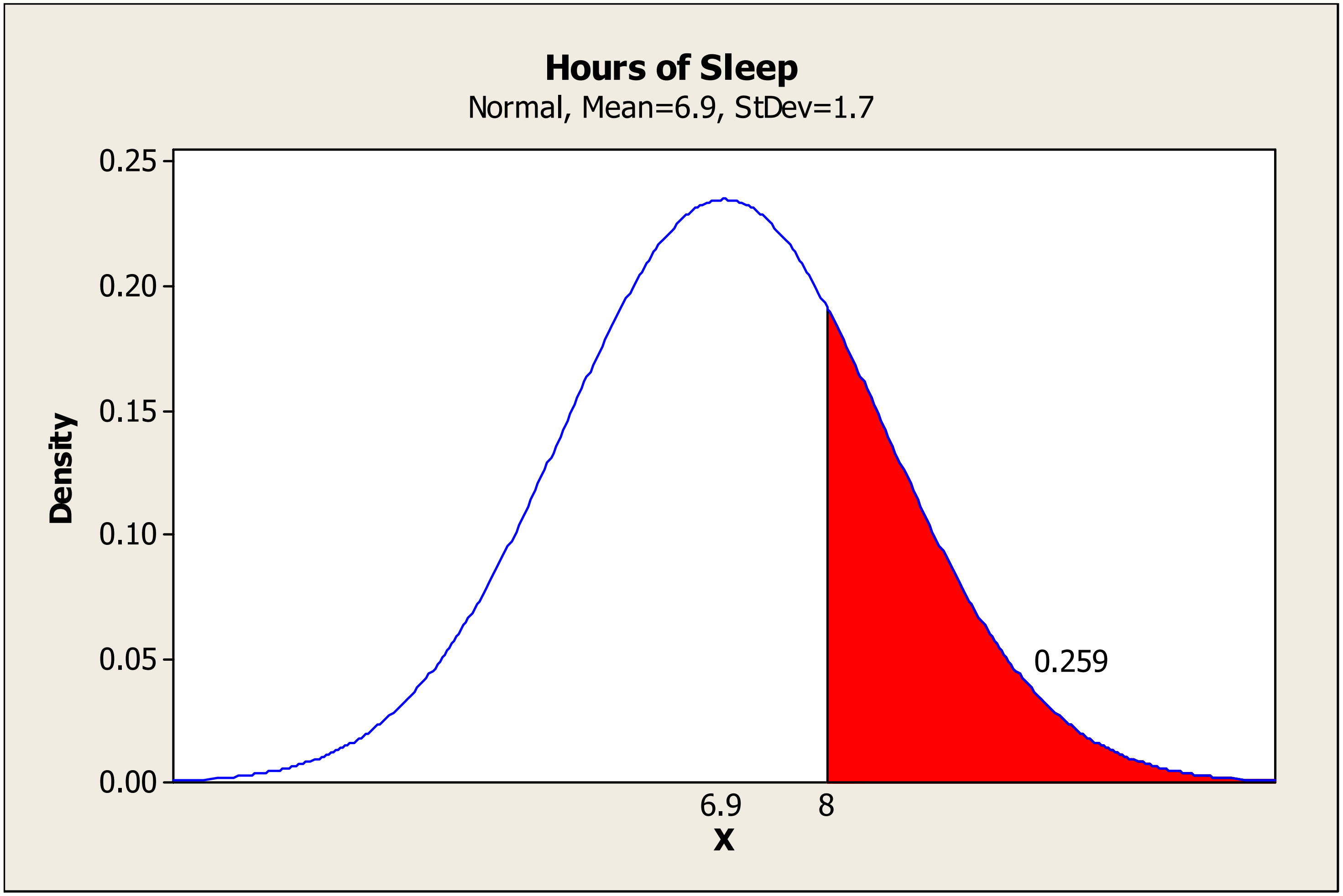
$P(z > +2.21)$

$P(-2.21 < z < 2.21)$

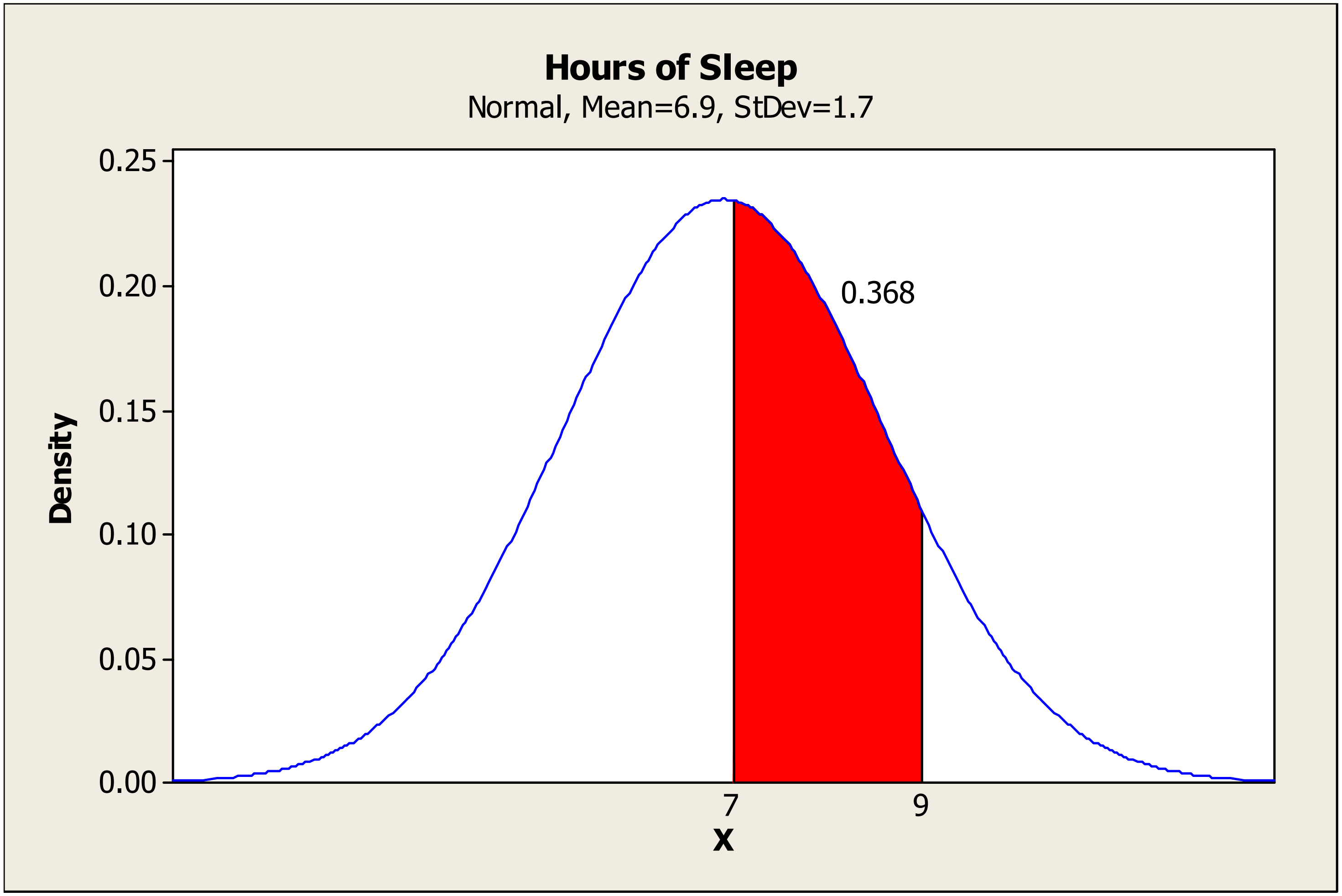
$P(-1.96 < z < 1.96)$

Some pictures for hours of sleep, mean = 6.9 and standard deviation = 1.7

Probability X > 8 = proportion who sleep more than 8 hours = .259



Probability 7 < X < 9 = proportion who sleep between 7 and 9 hours = .368



Some useful relationships for normal curve probabilities (a, b, d are numbers):
Illustrate on board, with examples:

$P(X > a) = 1 - P(X \le a)$
$P(a < X < b) = P(X \le b) - P(X \le a)$
$P(X > \mu + d) = P(X < \mu - d)$
$P(X < \mu) = .5$

Using R Commander:
Distributions → Continuous distributions → Normal distribution → Normal probabilities
Enter variable value, mu, sigma, then choose lower tail or upper tail.
Result shown in output window.

<u>Example</u>: For sleep hours, with mean mu = 6.9 and sigma = 1.7.
1. What proportion of students sleep more than 8 hours?
Use value = 8, mu = 6.9, sigma = 1.7, upper tail.
Result: 0.2587969 (about 26%)
2. What proportion of students get the recommended 7 to 9 hours of sleep?
Picture showed that it was about .368, or 36.8%. Show calculation on board.

Working backwards: We want to know the cutoff for a certain proportion

Example: What z-value has 95% of the standard normal curve below it?

Method 1: Use Table A.1. Find .9500 in the body of the table, then read z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
…										
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

Result: z = 1.645
What is the amount of sleep that only 5% of students exceed? (On board)
In general, $X = z\sigma + \mu$

Using R Commander:
Distributions → Continuous distributions → Normal distribution → Normal quantiles
Enter proportion of interest, mean, standard deviation, and upper or lower tail.
Ex: Height with 30% of women *above* it. Enter .3, 65, 2.7, upper. Get 66.41588.

NORMAL APPROXIMATION FOR BINOMIAL RANDOM VARIABLE

If X is a binomial random variable with n trials and success probability p, and if n is large enough so that np and n(1-p) are both at least 5 (better if at least 10), then X is *approximately a normal random variable* with:

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Therefore

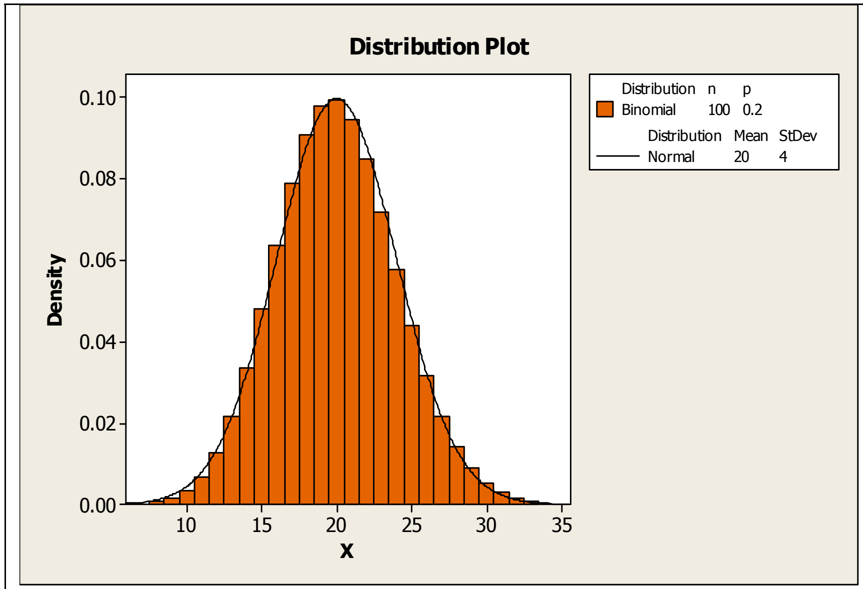
$$P(X \le k) \approx P(z \le \frac{k - np}{\sqrt{np(1-p)}})$$

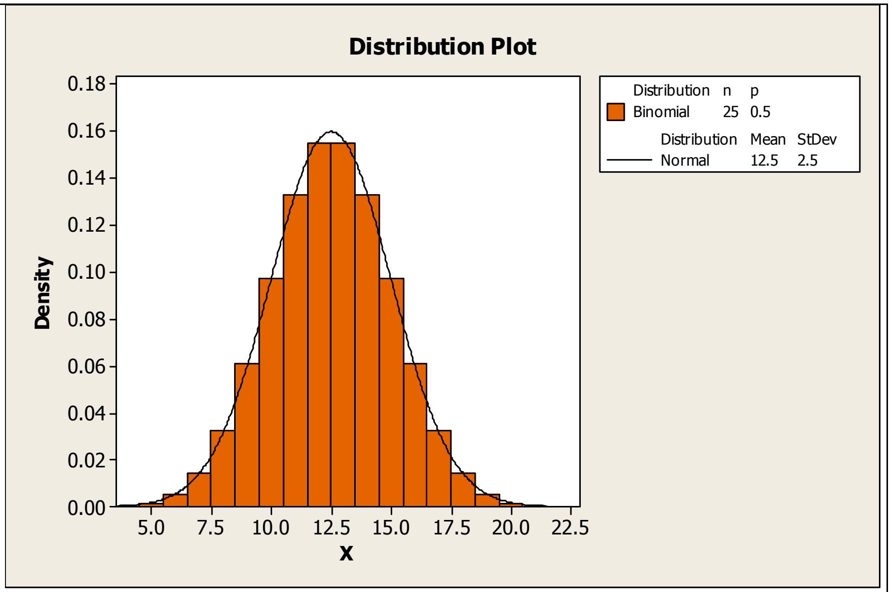
In other words, these are *almost* equivalent:

- Adding probabilities for all values from 0 to *k* for binomial with *n, p*
- Finding area under curve to the left of *k* for normal with μ = np, $\sigma = \sqrt{np(1-p)}$

Comparing the two distributions for some values of n and p:

n = 100, p = .2; $\mu = np = 20$, $\sigma = 4$ n = 25, p = .5; $\mu = np = 12.5$, $\sigma = 2.5$





Shaded rectangles show the binomial probabilities for each value on the x axis; smooth bell-shaped curves show the normal distribution with the same mean and standard deviation as the binomial.

Example from last lecture, where we found exact binomial probability:

A poll samples 1000 people from a population with 48% who have a certain opinion. X = number in the *sample* who have that opinion. What is the probability that a *majority* (at least 500) of the *sample* have that opinion?

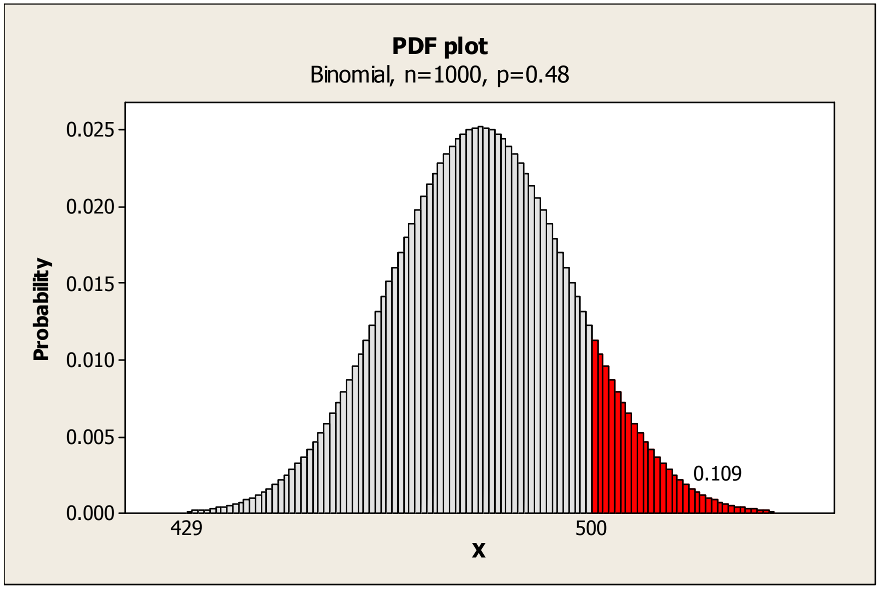
Binomial with n = 1000 and p = .48, μ = 480 and $\sigma = \sqrt{1000(.48)(.52)} = 15.8$

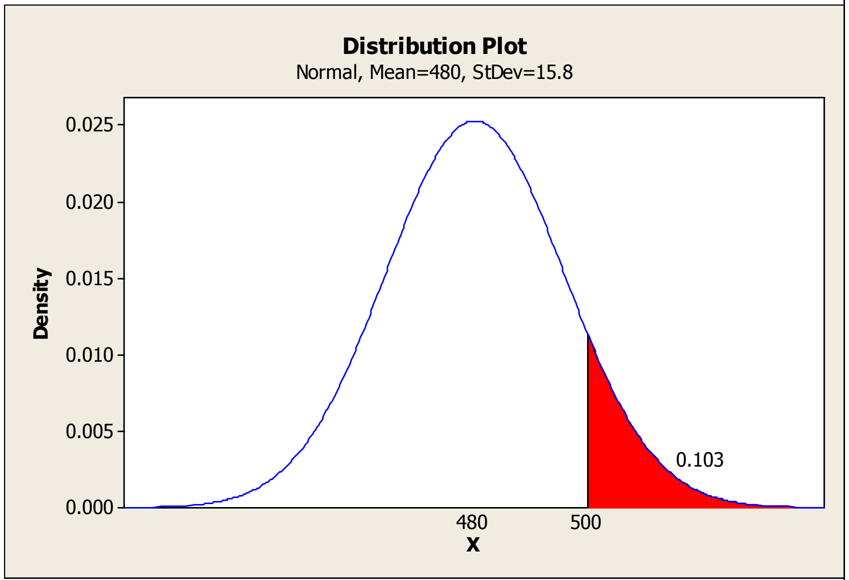
$$P(X \geq 500) \approx P(z \geq \frac{500-480}{15.8}) = P(z \geq 1.2658) = .103$$

So there is about a 10% chance that a *majority* of the sample have the opinion. Much easier than adding up P(X=500) + P(X=501) + ….. + P(X=1000).

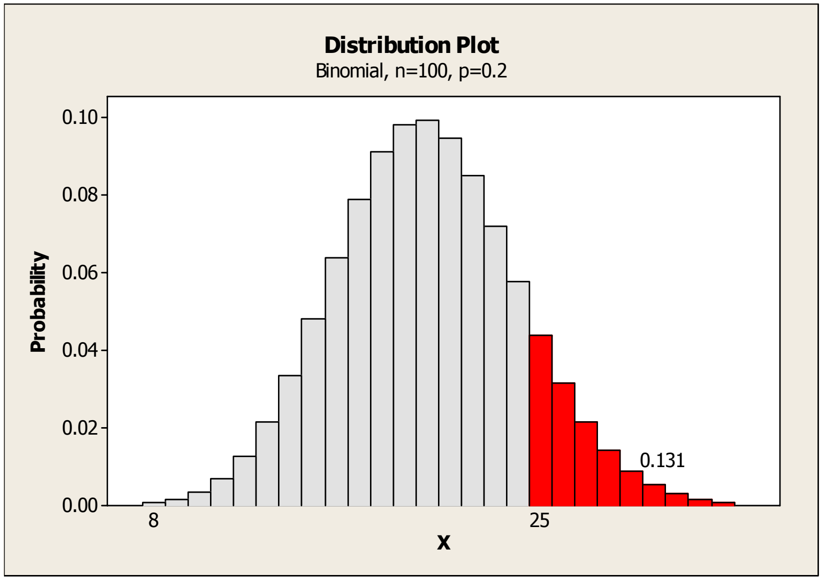
Picture on next page….

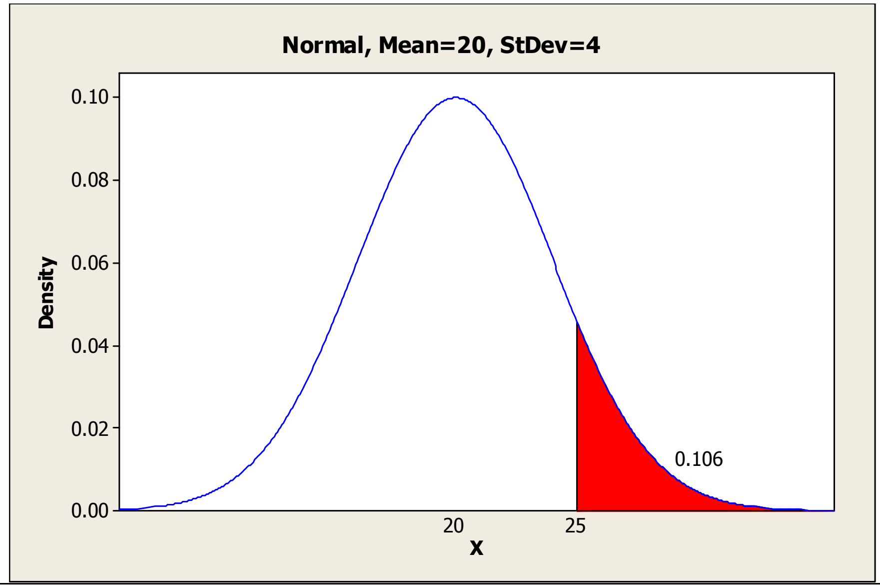
Comparing exact binomial and normal approximation: n = 1000 and p = .48





Example with smaller n: n = 100, p = .2 so μ = 20 and σ = 4





One final note about this: A more accurate place to start is either 0.5 above or below k, depending on the desired probability. Note that binomial starts at 24.5.

Ex: For n = 100 and p = .2, what is the probability of at least 25 successes?
Find P(X > 24.5) for normal X with μ = 20 and σ = 4.
Exact binomial probability of at least 25 successes is **0.1313**.
Normal P(X > 25) = 0.1056, but P(X > **24.5**) = **0.1303**.