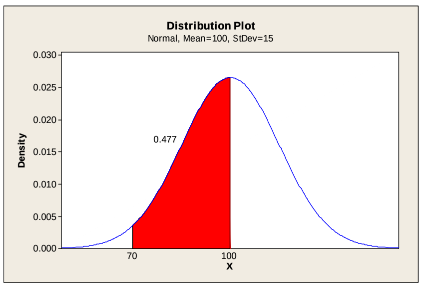
# Chapter 18: Sampling Distribution Models

## Sampling Distribution for the Sample Mean

1. $E(\bar{y}) = \mu$ $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$

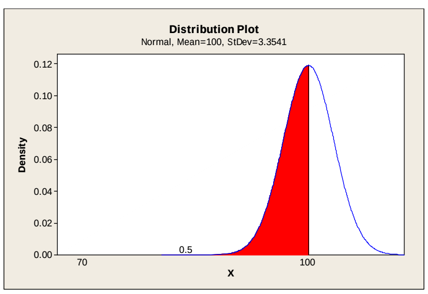
2. As you increase the sample size,
   a) the sampling distribution of the sample mean becomes MORE NORMAL.
   b) does $SD(\bar{y})$ increase or decrease? DECREASE
   c) Therefore, do individual values of $\bar{y}$ get closer or further from $\mu$, the theoretical mean of the distribution?
   CLOSER

3. IQ scores are believed to follow the normal model with mean 100 and standard deviation 15.
   a) What fraction of people have IQ scores between 70 and 100?




0.477

   b) What's the probability that the mean IQ of 20 people is between 70 and 100?



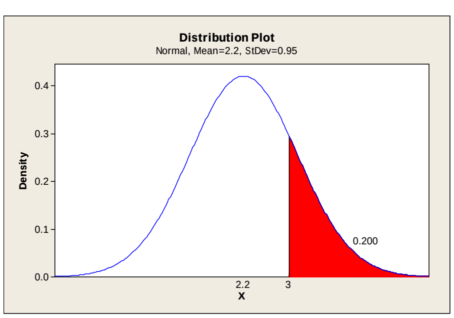

NEW STANDARD DEVIATION = 15 / SQRT(20) = 3.354
NEW MODEL: N(100, 3.354)

ANSWER: 0.5

4. The average sales price for a home in Beverly Hills was $2.2 million. Assume these prices have a standard deviation of 9.5 million. (Nov 07-Jan 08: http://www.trulia.com/home_prices/California/Los_Angeles-heat_map/ )
   a) Why is it unreasonable to assume that these homes are normally distributed?
   HOUSING PRICES ARE USUALLY RIGHT SKEW – WITH A FEW EXCEPTIONALLY PRICEY ONES (COMPARED TO THE REST IN THE NEIGHBORHOOD)

   b) Explain why we cannot determine that a given home sold for more than $3 million.
   WE DO NOT HAVE THE PROBABILITY MODEL. NO MODEL OR DATA, NO METHOD FOR CALCULATING.

   c) Can you estimate the probability that the mean selling price of 100 randomly selected homes sold is more than $3 million? Explain (and find the probability).



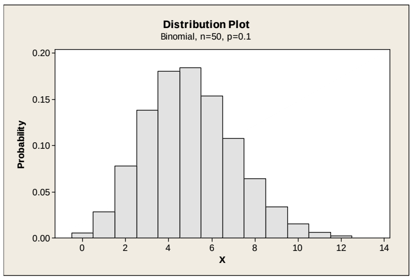

YES... THE CENTRAL LIMIT THEOREM APPLIES. THE MEAN PRICES WILL FOLLOW A NORMAL MODEL WITH MEAN 2.2 AND SD 9.5/SQRT(100) = 0.95... N(2.2, 0.95)

THEREFORE, THE PROBABILITY IS 0.200.

## Sample Distribution for Sample Proportion

1. Back to proportions… $E\left(\hat{p}\right) = p$ $\qquad$ $SD\left(\hat{p}\right) = \sqrt{\dfrac{pq}{n}}$

2. First, consider the binomial distribution from last time. It is claimed that 10% of M&Ms are green. Suppose that the candies are packaged in small bags containing about 50 M&M's. A class of elementary school students opens several bags, counts the various colors of the candies, and calculates the proportion that are green.
   a) If we plot a histogram of the proportions of green candies in the various bags, what shape would we expect it to have?




UNIMODAL, SLIGHTLY SKEW TO THE RIGHT
(IT'S ESSENTIALLY A SCALED BINOMIAL)

   b) Can that histogram be approximated by a Normal model?
      NO. IN ORDER TO APPLY THE NORMAL MODEL, I NEED TO EXPECT AT LEAST 10 SUCCESSES AND 10 FAILURES. BUT np = (50)(.1) = 5.

   c) What should the center of the histogram be?
      p = .1

   d) What should the standard deviation of proportion be?
      sqrt(pq/n) = sqrt((.1)(.9)/50) = 0.042

3. Same scenario as #4, but now the class buys bigger bags of candy, with 200 M&M's each.
   a) Explain why it's appropriate to use a Normal model to describe the distribution of the proportion of green M&M's they might expect.
      NOW np = (200)(.1) = 20 AND nq = (200)(.9) = 180
      THEREFORE, THE NORMAL MODEL APPLIES TO DESCRIBE THE DISTRIBUTION OF PROPORTION.
      MEAN = .1
      SD = SQRT(pq/n) = 0.021

   b) Use the 68-95-99.7 Rule to describe how this proportion might vary from bag to bag. (In particular, specify where the central 68% of the bags lie, etc…)
      68% OF THE VALUES LIE WITHIN 1 STANDARD DEVIATION, SO BETWEEN 0.079 AND 0.121
      95% OF THE VALUES LIE WITHIN 2 SD, SO BETWEEN 0.058 AND 0.142
      99.7% OF THE VALUES LIE WITHIN 3 SD, SO BETWEEN 0.037 AND 0.163

      IT WOULD BE **VERY** UNUSUAL FOR US TO SEE VALUES OUTSIDE 3.7% AND 16.3%.

   c) How would the model change if the bags contained even more candies.
      THE STANDARD DEVIATION WOULD GET SMALLER. THIS WOULD YIELD AN EVEN HIGHER CONCENTRATION OF THE SAMPLE PROPORTIONS AROUND THE CENTER.