Chapter 6: Normal Probability Distributions

§6.1 Review & Preview

In this chapter we will be discussing **continuous** probability distributions and hence **continuous random variables**. Recall the distinction between discrete and continuous quantitative data discussed in Chapter 1. We will be applying that distinction in this chapter. A continuous random variable has **infinitely many values** associated with a continuous scale.

This chapter will mainly be concerned with the continuous probability distribution known as the **Normal Distribution**. Here are the reasons that an entire chapter will be devoted to one distribution:
  1)   Much real life data conforms to this type of distribution
  2)   Important inferential statistics are based upon having normal data

The graph of the **Normal Distribution** is symmetric and bell-shaped. We have discussed this type of distribution in both chapters 2 & 3. The random variable, x, is continuous and the distribution can be described by the following formula, known as a **density function**.

$$f(x) \; = \; \frac{e^{-\frac{1}{2}\left[(x-\mu)/\sigma\right]^2}}{\sigma\sqrt{2\pi}} \qquad\qquad \begin{array}{l}\mu = \text{mean} \\ \sigma = \text{std. dev.}\end{array}$$

The graph of this distribution has the following characteristics:
  1)   It is bell-shaped
  2)   It is symmetric about the vertical line $x = \mu$
  3)   The x-axis forms a horizontal asymptote (it approaches, but never touches or crosses the x-axis)
  4)   The inflection points (point where the concavity changes sign) represent σ
  5)   Area under the curve is equal to 1
  6)   There is no area under the curve for a single value, $P(X=x)$ doesn't have any meaning on a continuous curve. We can only determine the area under the curve for an interval $P(X<x)$ or $P(X>x)$ or $P(x_s < X < x_B)$. This idea can get messy as we are really talking about a Calculus concept here.

Here is a picture of a normal curve. We'll use it to point out some of the key concepts we are discussing here.
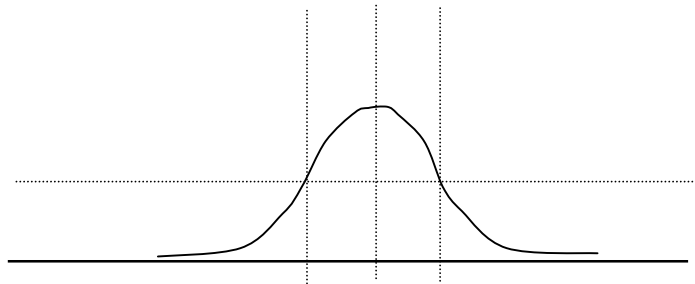
The normal density curve changes shape in two ways:
1)   Based on its location along the x-axis – change based upon μ
2)   Based on its width and height – change based upon σ

Let's look at some example using technology:

Example: Using your TI-83 or 84 we will plot 3 normal density curve with different mean.
Curve 1: $\mu = 0$ & $\sigma = 1$
Curve 2: $\mu = 1$ & $\sigma = 1$
Curve 3: $\mu = 2$ & $\sigma = 1$
a)   Find $\mu - 3.5\sigma$ for curve 1
b)   Find $\mu + 3.5\sigma$ for curve 3
c)   In Y= menu input the following for each curve, so that you have Y_1, Y_2 & Y_3 with equations for curve 1, 2 & 3
Input: 2nd VARS use normalpdf(X, μ, σ)
Window Menu: Input x-min a) & x-max b) values
                    Use y-min as 0 and y-max as 0.5
                    Scale each axis by 0.5

Example: You can repeat this experiment by leaving the mean constant and changing the standard deviation to see how the height of the curve will change each time
Curve 1: $\mu = 0$ & $\sigma = 1$
Curve 2: $\mu = 0$ & $\sigma = 2$
Curve 3: $\mu = 0$ & $\sigma = 3$
a)   Find $\mu - 3.5\sigma$ for curve 3
b)   Find $\mu + 3.5\sigma$ for curve 3
c)   In Y= menu input the following for each curve, so that you have Y_1, Y_2 & Y_3 with equations for curve 1, 2 & 3
Input: 2nd VARS use normalpdf(X, μ, σ)
Window Menu: Input x min a) & x max b) values
                    Use y min as 0 and y max as 0.5
                    Scale each x-axis by 1 & y-axis by 0.5

§6.2 The Standard Normal Distribution

Due to the fact that we are discussing *continuous probability distributions* we need to discuss the fact that the graph is still similar to what we see in a **probability histogram**. The area under the graph is still equal to 1 and the area under any section is equal to the section's width (value of the random variable) times its height (the probability of the occurrence of the given random variable). Whenwe talk about a graph of the probability function when our random variable is continuous, we call it a **density curve**.

Properties of a Density Curve
1) Area under the curve is equal to 1
2) Area of any section is equal to the probability of occurrence of the random variables describing the width of the section.
3) There is no area over any PARTICULAR value of a random variable, therefore **equality** DOES NOT matter. $P(X = x) = 0$ & $P(X \geq x) = P(X > x)$ are the same.

The easiest introduction to a continuous random variable we encounter is the **Continuous Uniform Distribution**. That is because the probability remains constant for the entire density function and the **density curve** is a RECTANGLE.

Example: The cycle time for trucks hauling concrete to a highway construction site is uniformly distributed over the interval 50 to 70 minutes.

a)   Draw the density curve. Label the probability. To do this recall that the area under the curve must be equal to 1. The curve has a length equal to the interval of minutes and therefore the height is simply an algebra problem $A = l \cdot w$.

b)   What is the average amount of time that it takes for a cycle? (What's the mean amount of time it takes?)

c)   What is the probability that it will take the truck more than 55 minutes? Use correct notation and show your computation of area under the curve.

d)   What is the probability that it will take the truck at most 65 minutes? Use correct notation and show your computation of area under the curve.

e)   What is the probability that it will take longer than 65 minutes if you know it's taken longer than 55 minutes? Hint: This is conditional probability from the definition that we learned in chapter 4.

f)   What's the probability that it will take **between** 55 and 65 minutes. Use correct notation and show your computation of area under the curve.

g)   What is the probability that it will take the truck at most 52 minutes? Use correct notation and show your computation of area under the curve.

The normal distribution yields probabilities in the same way that the continuous uniform does, but because the shape is more complicated it is harder to find the areas under the curve without more rigorous mathematics. Therefore we have a technique that makes finding these areas easier.

The normal distribution has many variations on the bell shape. We looked at some of these differences in the investigation using our TI calculators on page 3 of the notes. See above.

What we notice is:
1) Normal curves can move horizontally along the x-axis depending upon their mean.
2) Normal curves can be centered at the same position yet change in their height and spread due the variations. The larger the standard deviation the larger the spread and the lower the curve.

To find areas under the Normal Curve, we must repeatedly solve the probability density function. The Normal Density Function requires Calculus methods to solve. That is not within the realm of possibility for most users of the function so, the **Standard Normal Distribution** is used instead. The standard normal distribution has a **mean of zero** and a **standard deviation of one**. Since every normal density function is a transformation of the standard normal density function, we can use pre-calculated tables to find probabilities for any normal distribution through the standard normal. So, the first thing we want to learn is how to find probabilities for the Standard Normal Distribution.

The graph of the **Standard Normal Distribution** is symmetric and bell-shaped. We have discussed this type of distribution in both chapters 2 & 3. The random variable, x, is continuous and the distribution can be described by the following formula, known as a **density function**.

$$f(x) \ = \ \frac{e^{-1/2(x)^2}}{\sqrt{2\pi}} \qquad\qquad \mu = \text{mean} = 0$$

$\sigma = \text{std. dev.} = 1$

*Note: They "disappear"

All tables are setup to be read as cumulative from the left. As a result we must learn to setup the probability to read it in this manner.

When working with a left-tail table we can only find the area under the curve to the left of the random variable, X. Because of this limitation we must learn to rewrite probabilities in terms of being less than the random variable, X. Here are the short-cuts:

$P(X < x)$ Straight table look up.

$P(X > x) = 1 - P(X < x)$ [Technically it needs to $\le x$ but there is no area under x, so for any continuous distribution it doesn't matter.]

$P(x_s < X < x_B) = P(X < x_B) - P(X < x_s)$

Example: Let's start with a simple **straight** table look-up
a)   Find the probability of being less than -1 in a standard normal distribution.

b)   Find the probability of being less than 2.78 in a standard normal distribution.

c)   Find the probability of being less than 4.87 in a standard normal distribution.

Your Turn:
a)   Find the probability of being less than -1.3 in a standard normal distribution.

b)   Find the probability of being less than 1.52 in a standard normal distribution.

c)   Find the probability of being less than -3.51 in a standard normal distribution.

Now we will find probabilities. Since we have a TI-83/84 we can use the DIST menu and the normalcdf(left limit, right limit, mean, std dev) If you leave off the mean and the std dev the calculator will assume a standard normal, with mean, 0 and std dev, 1. For the left limit use a very small number (e.g. -1×10^{10})when looking for a left-tail, and if looking for a right-tail use a very large number (e.g. 1×10^{10}).

Example: Use the TI to find the probabilities above.
a)   Find the probability of being less than -1 in a standard normal distribution.

b)   Find the probability of being less than 2.78 in a standard normal distribution.

c)   Find the probability of being less than 4.87 in a standard normal distribution.

Your Turn:

a)   Find the probability of being less than -1.3 in a standard normal distribution.

b)   Find the probability of being less than 1.52 in a standard normal distribution.

c)   Find the probability of being less than -3.51 in a standard normal distribution.

Now we can do the more difficult scenarios – the larger than and the between values.

Example: Find the following probabilities. Use correct notation and draw a
picture representing the area being found under the standard
normal distribution.

a)   Find the probability of being more than -1.52

b)   Find the probability of being more than 2.3

c)   Find probability of being between 2 and 3.1

d)   Find the probability of being between -1.7 and 0.53

f)   Find the probability of being between -2.09 and -1.02

Our next goal is to learn how to find values for the random variable using the probability.
This is called the Inverse Normal or doing a reverse, table look-up. The critical values
that we use in statistical inference values from the inverse normal distribution.

As before, I'm going to teach you based upon doing it with a table, so that if you ever
find yourself without the TI-83/84 you could still do the problem.
       1)   Find the probability of being less than Z in the body of the table
       2)   Read to the "outside edges" to find the Z
              a)   Interpolation can be done to find a more exact value
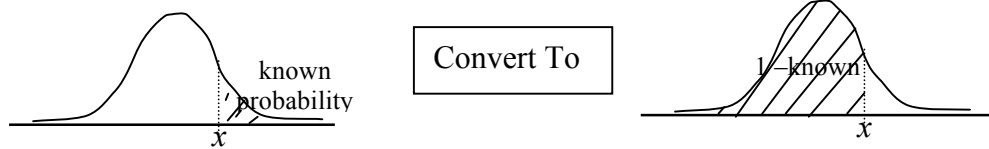   b)   Go to the closer Z value according to the probability

For the calculator, you will use DISTR menu again and this time
INVNORM(probability, μ,σ) Just as before, if you leave off mean and std dev, the
calculator will assume std. normal. You must realize that the probability is always the
probability of being less than a value. Unlike the NORMCDF() function, you will always
be finding values in terms of being less than a given value.

Just as before we will draw pictures to help us with the concepts.

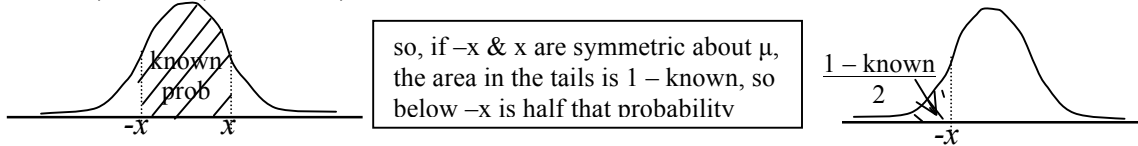1) $P(X < x)$ = known probability
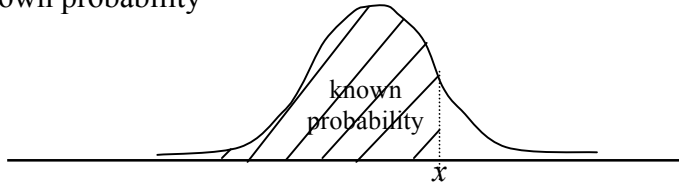
2) $P(X > x)$ = known probability

SO, $P(X > x) = 1 - P(X < x) =$ known $\therefore$ $P(X < x) = 1 -$ known

3) $P(-x < X < x) =$ known

so, if –x & x are symmetric about μ, the area in the tails is 1 – known, so below –x is half that probability

SO, $P(-x < X < x) = 1 - 2P(X < -x) =$ known $\therefore$ $P(X < -x) = \dfrac{1 - \text{known}}{2}$

Example: Use the Z-table to find the **z-score** representing the 99%tile. Recall that the percentile means the percentage of data lower than the given value.

*Note: The exact value does not exist so we must approximate or interpolate. We have 0.9898 and 0.9901 which are close to 0.9900, and since 0.9901 is closest we can use it's z as an approximation. Interpolation is more difficult, it requires finding the number of increments between 0.9898 and 0.9901 and then indicating what portion of the way 0.9900 is between them. If we divide the region up by 0.0001, making 0.9900 2/3 of the way between the two, so the z-score would be 2/3's of the way between 2.32 & 2.33, making it approximately 2.327.

Your Turn: Use the Z-table to find the **z-score** representing the 15th percentile.

Example: Use your calculator to find the values in the last 2 examples.

*Example: Critical values are values that represent cut-off points where (1 – probability)/2 will lie below or above the symmetric z-scores. They are notated as a subscript that indicates the 1 – probability

Example: The standardized weight of a fawn from Mesa Verde Nat'l Park is 0.65. What is the probability that a fawn in Mesa Verde Nat'l Park will have a weigh with a z-score of 0.65?

Example: What percentage of fawns will weigh more than a fawn with a standardized weight of -1.91?

Example: What are the standardized weights that you expect to represent the 96.5%tile?

Example: Find the critical values that put 54% of the data in the middle of the distribution.

§6.3 Applications of Normal Distributions

This section discusses how to use the Standard Normal Table to find the probability of any Normally Distributed random variable despite the mean and standard deviation. Since we will be using the TI-83/84 to find normal probabilities. We will discuss how to find probabilities for any normal distribution using a standard normal, just in case you ever find yourself without a calculator that will calculate these probabilities.

1)   Translate the random variable to a z-score
2)   Look up in a left-tail table, translating if necessary (a pix will help)
   a) $P(X < x) = P(Z < z)$
   b) $P(X > x) = 1 - P(X < x) = 1 - P(Z < z)$
   c) $P(x_S < X < x_B) = P(X < x_B) - P(X < x_S)$
         $= P(Z < z_B) - P(Z < z_S)$

Recall: *For finding probabilities with the TI-83/84, use DISTR menu and normalcdf(lower bound, upper bound, mean, std. dev. Remember if you don't put in mean and std dev. the calculator will assume std. normal.*

Example: Let's do an example where we draw the pictures and prepare for the look up using a Z-score, then we'll use our calculator to find the probabilities with mean 0 and std dev 1 and compare it with the answer from the original mean and std. dev.

Porphyrin is a pigment in blood protoplasm and other body fluids that is significant in body energy and storage. Let x be a random variable that represents the number of milligrams of porphyrin per deciliter of blood. In health adults, x is approximately normally distributed with mean of 38 and std. dev. of 12 (*Diagnostic Tests with Nursing Implications*, edited by S. Loeb, Springhouse Press). This is problem #26 p. 269 of your text.

   a)   Write the probability that a randomly chosen individual will have a porphyrin level lower than 60 using probability notation.
   - Next, convert it to a probability using a Z .
   - Sketch the picture of the probability that you are trying to find.
   - Use your calculator to find the probability, using the z-score.
   - Compare it to the answer for the probability of x under the original mean and std. dev.

b)   Answer all the same questions as a) for the
       probability that a randomly chosen individual will
       have a porphyrin level above 16.

c)   Answer all the same questions as a) for the
       probability that a randomly chosen individual will
       have a porphyrin level between 16 and 60.

d)   Answer all the same questions as a) for the
       probability that a randomly chosen individual will
   have a porphyrin level greater than 60.

Our next topic is how to find the value of the random variable given the probability of
that random variable's occurrence. As before, I'm going to teach you based upon doing
it with a table, so that if you ever find yourself without the TI-83/84 you could still do the
problem.
    1)   Find the z based upon the reverse table look up (look in the body)
    2)   Use $x = z\sigma + \mu$ to find x, since $z = \dfrac{x - \mu}{\sigma}$
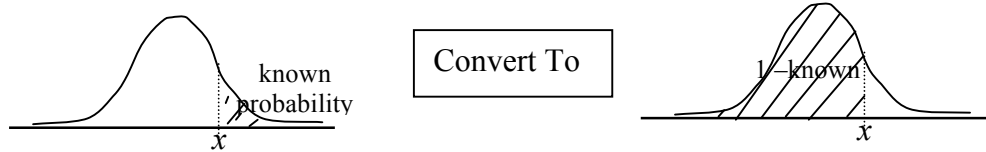
Since you will have your calculator, we'll practice all but the table look-up. You will use
DISTR menu again and this time INVNORM(probability, μ,σ) Just as before, if you
leave off mean and std dev, the calculator will assume std. normal.

Just as before we will draw pictures to help us with the concepts.
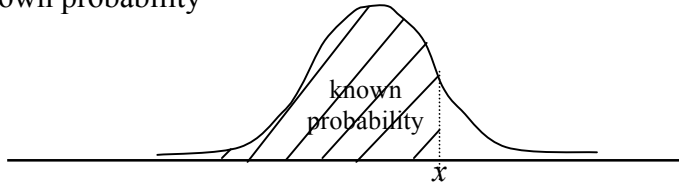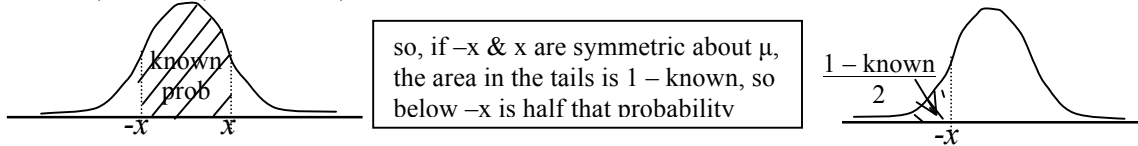
1) $P(X < x)$ = known probability

2) $P(X > x)$ = known probability

known probability

Convert To

$1 - $ known

SO, $P(X > x) = 1 - P(X < x) =$ known $\therefore$ $P(X < x) = 1 -$ known

3) $P(-x < X < x) =$ known

so, if $-x$ & x are symmetric about μ, the area in the tails is $1 -$ known, so below $-x$ is half that probability

$\dfrac{1 - \text{known}}{2}$

SO, $P(-x < X < x) = 1 - 2P(X < -x) =$ known $\therefore$ $P(X < -x) = \dfrac{1 - \text{known}}{2}$

Example: Human gestations periods are approximately normally distributed with a mean of 268 days and std. dev. of 15 days. Premature babies are those that are in the lowest 4%. Find the length of gestation, in days, that separates the premature babies from the non-premature babies.

Example: The number of years before replacement is necessary for a specific brand of CD players is approximately normally distributed with a mean of 9.4 years and std. dev. of 4.2 years. Find the length of time, in years, that represents the cut-off points for the amount of time that these CD players "consistently" last.
Recall that your book terms "consistently" within _____ std. dev. See p.11 Ch. 5 notes!

Example: Most exhibition shows open in the morning and close in the late evening. A study of Saturday arrival times showed that t he average arrival time was 3 hours and 48 minutes after the doors opened, and the std. dev. was estimated at about 52 minutes. Assume the arrival times have a normal distribution.
Remember units must be consistent! This is problem #37 from p. 270 of our text.

*a)   At what time, after the doors open, will all but 10% of the people who are coming to the Sat. show have arrived?

b)   At what time, after the doors open, will only 15% of the people who are coming to the Sat. show have arrived?

*c)   What are the arrival times that represent the cut-off points for which the "usual" amount of people will have arrived?   See p. 11 of Ch. 5 notes!

Note: I have changed the problem or added the problem to the original.

§6.4 Sampling Distributions

Recall from our studies in the Chapter 1, that a **sample** produces **statistics** by which we will estimate **parameters** of the **population**.

Your book discusses this in terms of the types of inferences that we will be making about populations in the coming chapters. Brase and Brase, further iterate that sample statistics are used to make **inferences** about population parameters due to constraints on time, money and effort. In reality, it is nearly an impossible task to summarize an entire population, and therefore to find true population parameters. Here is a list and description of the types of inferences we will be making in the coming chapters:

---

3 Types of Inferences

1)   **Estimation** which *estimates* the value of a parameter using sample statistics.

2)   **Testing** to formulate a *decision* about a population parameter. The decision will be based on sample statistics.

3)   **Regression** which will *predict/forecast* the results of a statistical variable.

In order to evaluate the reliability of the inferences being made relies on the distribution of the statistic being used. The distribution of the statistic is called a **sampling distribution**. This section introduces us to the idea that statistics have distributions of their own!

A **sampling distribution** is a probability distribution of all possible simple random samples of size n, from the same population.

*Note: The simple random samples must all be of the **same** size, n!*

The sampling distributions that are typically studied in introductory statistics classes are the distributions of:

x-bar	which are approximately Normally distributed	We'll study in 7.2 & 7.3
p-hat	which are also approximately Normally distributed	respectively
s^2	which is distributed with a distribution called Chi-Squared	
s_1^2/s_2^2	which is distributed with an F-distribution	

To see how a sample statistics takes on a distribution of it's own, let's take a look at some samples of waiting times, to the nearest whole minute, of people standing in line at a supermarket check out line.

> **Example:** These random samples were created using EXCEL and the values are from a distribution called the Poisson. The Poisson is the distribution that describes waiting times, and is the reason I used this distribution to create this random data. (The lambda that I used was 2. If you would like to know more about the Poisson distribution you are welcome to research it on your own.) What you'll notice based on a frequency histogram is that the average waiting time for samples of individual waiting times is approximately bell-shaped.

Something that also needs to be discussed is a concept that I mentioned in a discussion in Chapter 1, a sample statistic being an **unbiased estimator**. This means that the mean of the sample statistics distribution is equivalent to the value of the population parameter being estimated. This is true for several of the sample statistics that we will be studying and have studied. Two examples are: x-bar and p-hat

The second discussion is one about the **variability** of the sampling distribution. Sampling method and sample size both effect variability. As the sample size increases variability decreases, since the variance (measure of variability) decreases in relation to the original variability and the sample size (variance of sampling distribution is variance of the population divided by sample size, and the std. dev. is the square root of that). We saw this in part d) of Example 3 above. Because the variability about the original mean decreases as the number of months increased that meant that the probability increased as the number of months increased, because the standard deviation is decreasing. Draw a picture to show this and you will see.

prob based on factor of z

$-(\sqrt{n})\cdot z \qquad (\sqrt{n})\cdot z$

prob based on factor of z

$-(\sqrt{n})\cdot z \qquad (\sqrt{n})\cdot z$

Let's look at another example that will highlight the change in variability as sample size increases. As we, do this example, let's keep in mind that as sample size increases, the distribution becomes narrower, but there is more probability of being closer to the mean. Think of it in terms of z-score and it will make more sense, as was shown in the diagrams above.

Example 4:   Exercise #10 on p. 307 of Brase & Brase's *Understandable Statistics*, Ed. 9.

Suppose x's distribution has a mean of 5. Consider two corresponding distributions of x-bar. The first based on samples of size n = 49 and the second based on samples of size n = 81.

a)   What is the value of the mean of each of the two distributions of x-bar?

b)   For which x-bar distribution is P(x-bar>6) smaller? Explain.

c)   For which x-bar distribution is P(4 < x-bar < 6) greater? Explain.

§6.5 Central Limit Theorem

The CLT (Central Limit Theorem) says that as the number of samples from any population increases their distribution will approach a normal distribution with mean equal to the mean of the original sample and a standard deviation of the original divided by the $\sqrt{n}$.

Central Limit Theorem
X is a random variable from any distribution with mean μ and
    standard deviation σ.
n samples are drawn
The distribution of the means of the samples (x-bars) will be $N\sim(\mu_{x\text{-}bar}, \sigma_{x\text{-}bar})$

$$\mu_{x\text{-}bar} = \mu$$
$$\sigma_{x\text{-}bar} = \sigma/\sqrt{n}$$

*The difference between this section and section 5.3 is the difference between a single sample and a bunch of samples. In section 5.3 we were dealing with a single sample with a mean of x-bar and a std. dev. of s. In this section we are dealing with many samples and the average of the samples' means.

Example 1:   A soft drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 mL and a standard deviation of 15 mL. What is the probability that

a)   A single serving has a t least 204 mL (assume normal distributed r.v.)

b)   the average amount dispensed in a random sample of 36 is at least 204 mL

*Note: Part a) is as it was in section 5.3 and part b) is using the CLT.

Example 2:   A random sample of size 64 is taken from a normal population with mean of 51.5 and std. dev. of 6.8. What is the probability that the sample mean will
a)   exceed 52.9?

b)   fall between 50.5 and 52.3?

c)   be less than 50.6?

d)   what's the probability that a single observation would be below 50.6?

Example 2:   A random sample of size 64 is taken from a normal population
with mean of 51.5 and std. dev. of 6.8. What is the probability that
the sample mean will

   a)   exceed 52.9?

   b)   fall between 50.5 and 52.3?

   c)   be less than 50.6?

   d)   what's the probability that a single observation would be below
        50.6?

Example 3:   We're going to do #18 on p. 309 of Brase&Brase's
Understandable Statistics, Ed. 9.

   Dean Witter European Growth is a mutual fund that specializes in stocks
from the British Isles, continental Europe and Scandinavia. The fund has over 100
stocks. Let x be a random variable that represents the monthly percentage return from
this fund. Based on information from *Morningstar Guide to Mutual Funds*, x has a mean
of 1.4% and standard deviation of 0.8%.

   a)   Is the average monthly return on the Dean Witter European
        Growth fund approximately, normally distributed? Explain.

   b)   After 9 months, what is the probability that the average montly
        percentage return (an average of the 9 months' x's) will be between 1%
        and 2%?

   c)   After 18 months, what is the probability that the average monthly
        percentage is between 1% and 2%?

d) Compare the probability in b & c. Which is larger? Does this make sense? (Think back to our soda example. As there are more sodas to average, doesn't the variation on the soda narrow, and therefore, the probability for one soda to be a particular value is not as high as a sample of 5 and that isn't as high as a sample of 30 and that isn't as high as a sample of 100.) Why would this happen?

e) At the of 18 months, the average monthly percentage return is more than 2%. Would that shake your confidence that the population mean is 1.4%? If after 18 months, the average was more than 2%, would that mean that the European markets were "heating up"? Explain your answer in terms of probability.

§6.6 Normal Approximation to the Binomial

This approximation is used appropriately when the following conditions are met.

np ≥ 5 and nq ≥ 5   **or**   n > 30

If   X ~ Binomial(n,p,q) then X ~ N(μ, σ) where
μ = np
σ = √npq

The only trouble with the normal approximation to the binomial is the discrepancy between the types of distributions. The binomial is discrete. Remember the "chunky" probability histograms that we used to show the shape of the distribution? Well, this causes some problems because there is area under the curve for a given value in the binomial, whereas in a continuous distribution such as the normal, there is no area under the curve for any particular value. As such, we have to "fake" the area. This is done with a **continuity correction**. This takes into account that each bar on the binomial is one unit wide, so for each value we subtract 0.5 and add 0.5 to the value to approximate that interval. Below is what it will look like on the normal curve.
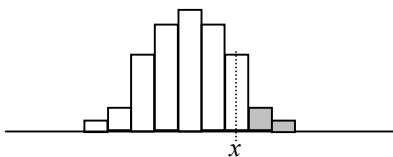
P(X = x) where x is Binomially Dist
➔ P(x − 0.5 < X < x + 0.5) where x is Normal with μ = np & σ = √npq
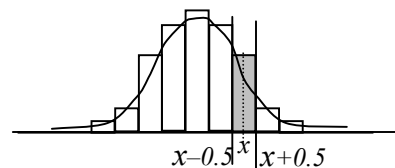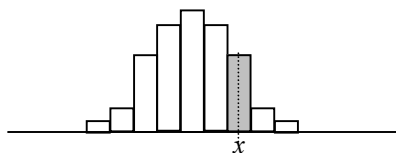
x

x−0.5 *x* *x+0.5*

P(X > x) where x is Binomially Dist
➔ P(X > x + 0.5) where x is Normal with μ = np & σ = √npq

x

x *x+0.5*

P(X ≥ x) where x is Binomially Dist
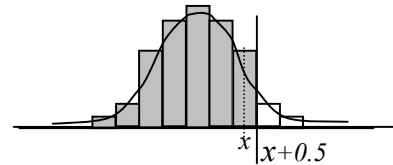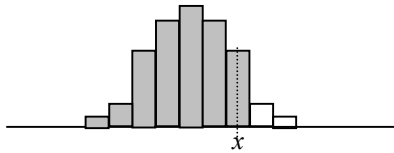➔ P(X > x − 0.5) where x is Normal with μ = np & σ = √npq

x

x−0.5 *x*
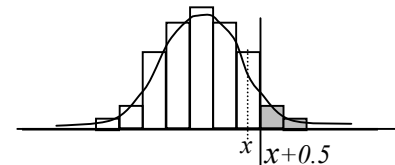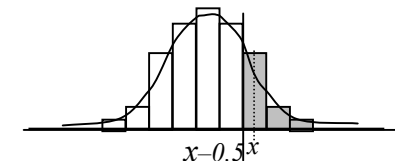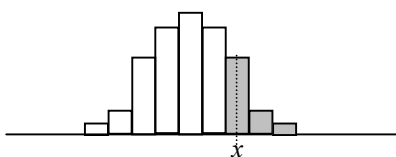
P(X < x) where x is Binomially Dist
  → P(X < x – 0.5) where x is Normal with μ = np & σ = √npq
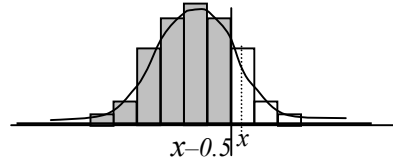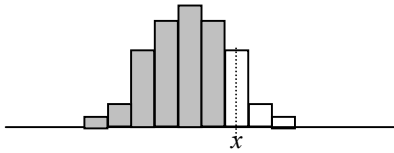
P(X ≤ x) where x is Binomially Dist
  → P(X < x + 0.5) where x is Normal with μ = np & σ = √npq

Now, all we need to do is try some examples!!

Example: *Use the normal approximation to the binomial to find the probability that at least 70 of 100 mosquitoes will be killed with an insect spray when the probability of killing them with the spray is 0.75. Use your calculator to calculate the actual probability too and compare.

Step 1:   What are n, p & x?

Step 2:   Calculate μ&σ

Step 3:   Define the probability in terms of the binomial & do the continuity correction.

Step 4:   Find the probability using the normal distribution

Your Turn: *If 23% of all patients with high blood pressure have bad side effects from a certain medicine, use the normal approximation to find the probability that among 120 patients more than 32 will have bad side effects.

*Note: Both examples are from Wadpole p. 230 Ex. 25 & p. 238 Ex. 23.

§6.7 Assessing Normality

One of the questions that is of concern when analyzing data is whether we have data that is approximately normally distributed. When we can answer this question in the affirmative we have an easier time analyzing data, especially **small** data sets that don't conform to the requirements of the Central Limit Theorem for their normality. We are now going to see some ways of checking for Normality.

Approximately Normal?

1)   Histogram
         $\approx$ Bell Shaped (Symetric)

2)   Outliers
         Not > 1 ➔ Outliers below $Q_1 - 1.5IQR$ or beyond $Q_3 + 1.5IQR$

3)   Skewness
         Pearson's Index = $\dfrac{3(\text{x-bar} - \text{x-tilde})}{s}$ is between -1 & 1

4)   Normal Probability (Quantile) Plot
         Z-score (x's) plotted against values of R.V, x (y's) forms
                 $\approx$ straight line

Triola only discusses Normal Quantile Plots but I wanted to be thorough in my discussion and have therefore included some of the things that other books include as well.

Technology can help us with:
1)   Histograms as we have already seen
      BTW here are the instruction for doing a histogram with EXCEL's data analysis package. I had forgotten that a histogram was possible with EXCEL.
            Histogram
              Input Upper Class Limits into another column
              Tools→Data Analysis→Histogram
               Highlight the data in the 1st column of your original workbook
              Click on Bin Range and highlight the column containing upper class limits
              Uncheck the labels box below Bin Range
              Check New Workbook
              Check Chart Output
             Click on OK
2)   Pearson's Skewness: EXCEL &TI used to get mean, median & std. dev., but calculation must be done by hand
3)   Normal Probability Plot: Both EXCEL & the TI we can force the issue
      Enter raw scores & find the Z-Score for each
            In EXCEL =STANDARDIZE(x, μ, σ)
            Now, use a scatterplot with just dots and plot standardized values
            against the x values (make sure std. vals are in 1st column & x vals are in 2nd

column)

In TI's DATA EDITOR with the cursor on the L2, type in (L1-mean)/std dev, where L1 contains the data and L2 will contain the z-scores
Now, use the STAT PLOT menu and turn on the first one that looks like dots by putting your cursor on the graph and entering. In the X-List enter L1, and in the Y-List enter L2. Mark with the square. ZOOM and choose ZOOMSTAT.

Example: Now let's test the normality of the following data which represents the average number of murders per capita for a large city.
12.6, 9.5, 14.9, 11.9, 11.0, 10.7, 9.7, 10.5, 8.2, 7.8, 9.5, 13.1, 8.2, 7.1, 9.2, 8.3, 8.7, 9.6, 10.9, 9.4, 12.1, 15.6, 12.0

a) Create a histogram of the data
b) Calculate Pearson's Skewness
c) Are there any outliers?
f) Make a Normal Probability Plot