7.1 - Continuous Probability Distribution and The Normal Distribution

Since a continuous random variable $x$ can assume an infinite number of uncountable values, we have to look at $x$ assuming a value within an interval.

The probability distribution of a continuous random variable is often presented in the form of a probability density curve (also called probability distribution curve). The total area under the curve must equal 1.

The area under the curve between two values $a$ and $b$ has two interpretations:
1. It is the proportion of the population whose values are between $a$ and $b$.
2. It is the probability that a randomly selected individual will have a value between $a$ and $b$.

The probability that a continuous random variable $x$ assumes a value within a certain interval is given by the area under the curve between the two limits of the interval.

The probability that a continuous random variable $x$ assumes a single value is always zero. It follows from this that $P(a \leq x \leq b) = P(a < x < b)$.

![Shaded area is 1.0 or 100%](image1)

![Shaded area gives the probability $P(a \leq x \leq b)$](image2)

![Shaded area gives the probability $P(65 \leq x \leq 68)$ as well as $P(65 < x < 68)$](image3)

**Figure 6.7** The probability of a single value of $x$ is zero.

**Figure 6.8** Probability “from 65 to 68” and “between 65 and 68.”
A continuous random variable can have many different distributions. One of the most widely used distributions is the normal probability distribution.

A Normal Probability Distribution gives a bell-shaped curve, symmetric about the mean, and with the two tails of the curve extending indefinitely. The total area under the curve is 1.0.

The equation of the normal distribution curve is

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2}
\]

Luckily, we will never have to use this equation in our class.

The parameters of the normal distribution is _______ and ________ .

There is not just one normal distribution curve but a whole family of them, each one depending on \( \mu \) and \( \sigma \). A larger standard deviation gives a wider curve than a smaller standard deviation, while the mean decides where to center the curve.
The *standard normal distribution* is a normal distribution with $\mu = 0$ and $\sigma = 1$.

![Normal Distribution](image)

The random variable that has a standard normal distribution is denoted $z$. The units marked on the horizontal axis are denoted $z$ and are called the $z$-values.

A $z$-value gives the distance from the mean in terms of number of standard deviations. That is, $z = 2$ is located 2 standard deviations to the right of the mean (right because it is a positive number), and $z = -1.5$ is located 1.5 standard deviations to the left of the mean (left because it is a negative number).

We can look up probabilities for different $z$-values using a table or by using our calculator. In this class we will use our calculators. This is how to do it on a TI-83/84:

Select PRGM -> NORMAL83 -> [ENTER] -> [ENTER] -> choose the option that fits your specific problem and enter $\mu = 0$, $\sigma = 1$, as well as your limit(s) -> [ENTER]

We can also reverse the procedure and *find the $z$-value(s) corresponding to a known area* under the standard normal distribution curve rather than vice versa.

We can find $z$-values using our TI-83/84 calculator:

Select PRGM -> INVNOR83 -> [ENTER] -> [ENTER] -> choose the option that fits your specific problem and enter the known area.

<table>
<thead>
<tr>
<th>We Know</th>
<th>Calculator Program</th>
<th>We Get</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$-value</td>
<td>NORMAL83</td>
<td>Area / Probability</td>
</tr>
<tr>
<td>$Area / Probability$</td>
<td>INVNOR83</td>
<td>$z$-value</td>
</tr>
</tbody>
</table>

For the chapter 7 homework, as well as on any quizzes and exams, use the appropriate calculator program rather than the tables in the appendices to calculate any probabilities, but be sure to include what program you are using, the parameters, and an appropriate, labeled picture.
Find the following:

(a) the area under the normal curve from $z = .23$ to $z = 2.10$  

(e) $P(z \geq 7.65)$

(b) the area under the normal curve to the left of $z = -.45$  

(f) $P(z = -1.38)$

(c) $P(-2 \leq z \leq 2)$  

(g) the z-score for which the area to its right is 0.25.

(d) $P(z \geq -0.40)$  

(h) the z-scores that bound the middle 70% of the area under the normal curve.
7.2 - Applications of the Normal Distribution

If \( x \) is a value from a normal distribution with mean \( \mu \) and standard deviation \( \sigma \), then we can convert \( x \) to a \( z \)-score by using the formula

\[
    z = \frac{x - \mu}{\sigma}
\]

We say that we standardize \( x \) when using the above process. The \( z \)-score follows a standard normal distribution, so we can use the methods from 7.1 to find areas under any normal curve by first converting the original values to \( z \)-scores.

However, with our calculators help, we usually don’t need to convert \( x \)-values to \( z \)-scores first. When using the program Normal83 on a TI-83/84 calculator, we can just enter the mean and standard deviation for the particular normal curve that we are working with and it will find the area/probability between some given \( x \)-values.

When you use the InvNor83, it will first give you the \( z \)-value(s), but you then have the choice of continuing to receive corresponding \( x \)-values by entering the mean and standard deviation.

<table>
<thead>
<tr>
<th>We Know</th>
<th>Calculator Program</th>
<th>We Get</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )- or ( z )-value</td>
<td>NORMAL83</td>
<td>Area / Probability</td>
</tr>
<tr>
<td>Area / Probability</td>
<td>INVNR83</td>
<td>( x )- or ( z )-value</td>
</tr>
</tbody>
</table>

1. Let \( x \) be a continuous random variable that follows a normal distribution with a mean of 550 and a standard deviation of 75.

(a) Find \( P(x > 600) \)

(b) Find \( P(410 < x < 510) \)
(c) Find the value of $x$ so that the area under the normal curve to the right of $x$ is 0.0275.

(d) Find the value of $x$ so that the area under the normal curve between $\mu$ and $x$ is approximately 0.4700 and the value of $x$ is less than $\mu$.

2. IQ scores are normally distributed with a mean of 100 and a standard deviation of 15.

(a) Find the probability that a randomly selected adult has an IQ greater than 131.5 (the requirement for membership in the Mensa organization). Is it unusual to have an IQ greater than 131.5?

(b) What percentage of people have IQ scores between 85 and 125?

(c) Find the IQ score separating the top 35% from the others.

(d) Find the IQ score that corresponds to the $z$-score -1.5. Interpret the result.
3. Rockingham Corporation makes electric shavers. The life (period before which a shaver does not need a major repair) of Model J795 of an electric shaver manufactured by this corporation has a normal distribution with a mean of 70 months and a standard deviation of 8 months.

(a) What percentage of shavers of this model has a life of at least 7 years?

(b) The company is to determine the warranty period for this shaver. Any shaver that needs a major repair during this warranty period will be replaced free by the company. What should the warranty period be if the company does not want to replace more than 1% of the shavers?

4. Engineers must consider the breadths of male heads when designing motorcycle helmets. Men have head breadths that are normally distributed with a mean of 6.0 in. and a standard deviation of 1.0 in. Due to financial restraints, the helmets will be designed to fit all men except those with head breadths that are in the smallest 2.5% or largest 2.5%. Find the minimum and maximum head breadths that the helmets will fit.
5. A machine at Keats Corporation fills 64-ounce detergent jugs. The machine can be adjusted to pour on average, any amount of detergent into these jugs. However, the machine does not pour exactly the same amount of detergent into each jug; it varies from jug to jug. It is known that the net amount of detergent poured into each jug has a normal distribution with a standard deviation of 0.35 ounce. The quality control inspector wants to adjust the machine such that at least 95% of the jugs have more than 64 ounces of detergent. What should the mean amount of detergent poured by this machine into these jugs be?

6. The weight of male babies less than 2 months old in the United States is normally distributed with mean 11.5 pounds and standard deviation 2.7 pounds. Find the 80th percentile of the male baby weights.

7. Speeds of automobiles on a certain stretch of freeway at 11:00 PM are normally distributed with mean 65 mph. Twenty percent of the cars are traveling at speeds between 55 and 65 mph. What percentage of the cars is going faster than 75 mph?
7.3 - Sampling Distribution and the Central Limit Theorem

A population parameter (ex. \( \mu, \sigma \)) is always constant, but a sample statistics (ex. \( \bar{x}, s \)) is always a random variable, because it will depend on what elements are included in the sample. That is, different samples from the same population can have different means for instance.

The **Sampling Distribution of** \( \bar{x} \) **is** the probability distribution of all possible values of \( \bar{x} \), when all possible samples of the same size \( n \) are taken from the same population.

There are \( \binom{N}{n} \) different samples of size \( n \) that we can pick from a population of size \( N \).

The mean and SD of the sampling distribution of \( \bar{x} \) are called the mean and standard deviation of \( \bar{x} \) and denoted by \( \mu_{\bar{x}} \) and \( \sigma_{\bar{x}} \) respectively.

\[
\mu_{\bar{x}} = \mu \\
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

The standard deviation \( \sigma_{\bar{x}} \) is sometimes called the **standard error** of the mean.

Note that \( \sigma_{\bar{x}} \) is smaller than \( \sigma \) and as the sample size increases, \( \sigma_{\bar{x}} \) decreases. That is, the sample means gets closer and closer to \( \mu \).

When the population from which samples are drawn is normally distributed, then the shape of the sampling distribution of \( \bar{x} \) is also normally distributed.

According to the **Central Limit Theorem**, the sampling distribution of \( \bar{x} \) is approximately normal for a large sample size, regardless of the shape of its population distribution.

The approximation becomes more accurate as the sample size increases.
A sample is generally considered large if \( n > 30 \).

The z-value for a value of \( \bar{x} \) is calculated as

\[
z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}
\]
Applications of the Central Limit Theorem

Using our calculators to find areas under the normal curve, we can use the central limit theorem to make statements as follows:

1. If we take all possible samples of the same (large) size from a population, then about 68.26% of the sample means will be within one standard deviation of the population mean.

2. If we take one large sample from a population, the probability that this sample mean will be within one standard deviation of the population mean is 0.6826.

This last statement is what we find more useful, since we in real life never look at ALL possible samples, but instead we want to select ONE sample and find the probability that the value of \( \bar{x} \) from this sample falls within a given interval.

ex. The print on the package of 100-watt General Electric soft-white light-bulbs says that these bulbs have an average life of 750 hours. Assume that the lives of all such bulbs have a normal distribution with a mean of 750 hours and a standard deviation of 55 hours. Find the probability that the mean life of a random sample of 25 such bulbs will be less than 725 hours.

ex. The annual per capita (average per person) chewing gum consumption in the United States is 200 pieces. Suppose that the standard deviation of per capita consumption of chewing gum is 145 pieces per year.

(a) Find the probability that the average annual chewing gum consumption of 84 randomly selected Americans is more than 220 pieces.
(b) Find the probability that the average annual chewing gum consumption of 84 randomly selected Americans is within 100 pieces of the population mean.

(c) Find the probability that the average annual chewing gum consumption of 16 randomly selected Americans is less than 100 pieces.

7.4 - Population and Sample Proportion

The Population Proportion, denoted by $p$, is the ratio of the number of elements in a population with a specific characteristic to the total number of elements in the population.

The Sample Proportion, denoted by $\hat{p}$ (read as "p hat"), is the same ratio but for a sample.

$$p = \frac{X}{N} \quad \text{and} \quad \hat{p} = \frac{x}{n}$$

Note that the relative frequency of a category or class gives the proportion that belongs to that category or class, and the probability of success in a binomial experiment also represents a proportion.

ex. In a random sample of 1000 subjects, 640 possess a certain characteristic. A sample of 40 subjects selected from this population has 24 subjects who possess the same characteristic. What are the values of the population proportion and sample proportion?
\( \hat{p} \) is a random variable (since it will vary depending on which elements are included in the sample), thus it has a probability distribution.

The **Sampling Distribution of the Sample Proportion**, \( \hat{p} \), is the probability distribution of \( \hat{p} \), which gives all different values that \( \hat{p} \) can assume and their probabilities.

The mean and standard deviation of the sample proportion, \( \hat{p} \), is denoted by \( \mu_\hat{p} \), and \( \sigma_\hat{p} \) respectively.

\[
\mu_\hat{p} = p \quad \text{and} \quad \sigma_\hat{p} = \sqrt{\frac{p(1-p)}{n}}
\]

According to the **Central Limit Theorem for Proportions**, the sampling distribution of \( \hat{p} \) is approximately normal for a large sample size.

The approximation becomes more accurate as the sample size increases.

A sample is generally considered large if \( np \geq 10 \) and \( n(1-p) \geq 10 \).

ex. Gluten sensitivity affects approximately 15% of people in the US. Let \( \hat{p} \) be the proportion in a random sample of 800 individuals who have gluten sensitivity. Find the probability that the value of \( \hat{p} \) is

(a) within 0.02 of the population proportion

(b) not within 0.02 of the population proportion

(c) greater than the population proportion by 0.025 or more
Seventy percent of adults favor some kind of government control on the prices of medicines. What is the probability that the proportion of adults in a random sample of 400 who favor some kind of government control is

(a) less than 0.65

(b) between 0.73 and 0.76

(c) within 0.06 of the population proportion
7.6 – Assessing Normality

When a sample is large, we have the Central Limit theorem to ensure that $\bar{x}$ is normally distributed. But when the sample is small we need to determine if the population, from which the sample was taken, is approximately normal in order to be able to use the techniques learnt in chapter 7.

We will NOT assume a population to be normal if

- The sample contains an outlier
- The sample exhibits a large degree of skewness
- The sample is multimodal

If a sample has none of these features, we can treat the population as a being approximately normal.

We can use dotplots, boxplots, stem-and-leaf plots, and histograms to help us detect above features.

Determine if it’s reasonable to treat the following as samples from approximately normal populations:

ex.

- [Graph of a dotplot with data points at 65, 70, 75, 80, 85, 90, 95, 100.]

ex.

- [Boxplot with an outlier at 2500, and a data point at 5000.]

ex.

- [Stem-and-leaf plot with data: 2|3, 3|7, 4|113, 4|55889, 5|223444, 5|66, 6|3, 6|6.]

ex.

- [Histogram showing frequency of diameters.]


A more sophisticated way to assess normality is to use **Normal Quantile Plots**, also called **Normal Probability Plots**.

Below is an example of the distribution of systolic blood pressure, for a random group of healthy patients.

![Histogram and Normal Quantile Plot](image)

Looking at the histogram, you can see the sample is approximately normally distributed. But the bar heights for 120-122 and 122-124 make the distribution look slightly skewed, so it’s not perfectly clear.

The normal quantile plot to the right is clearer. It shows the *observations* on the X axis plotted against the *expected normal score* (Z-score) on the Y axis. It's not necessary to understand what an *expected normal score* is, nor how it's calculated, to interpret the plot. All you need to do is check that the points roughly follow a straight line. If the points roughly follow a line – as they do in this case – the sample has a normal distribution.

The steps for constructing a normal quantile plot on the TI-84 PLUS Calculator are:

1. **Step 1.** Enter the data into **L1** in the data editor.
2. **Step 2.** Press **2nd**, **Y=** to access the STAT PLOTS menu and select Plot1 by pressing **1**.
3. **Step 3.** Select **On** and the normal quantile plot icon.
4. **Step 4.** For **Data List**, select **L1**, and for **Data Axis**, choose the **X** option.
5. **Step 5.** Press **ZOOM** and then **9: ZoomStat**.

**ex.** A placement exam is given to each entering freshman at a large university. A simple random sample of 20 exam scores is drawn, with the following results.

<table>
<thead>
<tr>
<th>61</th>
<th>60</th>
<th>60</th>
<th>68</th>
<th>63</th>
<th>63</th>
<th>94</th>
<th>66</th>
<th>65</th>
<th>98</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>71</td>
<td>74</td>
<td>63</td>
<td>66</td>
<td>61</td>
<td>61</td>
<td>65</td>
<td>72</td>
<td>85</td>
</tr>
</tbody>
</table>

*Construct a normal probability plot to determine if the exam scores are approximately normal.*