Normal Distribution Notes

**Population**: all measurements or observations of interest
**Sample**: a subset of the population

<table>
<thead>
<tr>
<th>Notation</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>$\bar{x}$</td>
<td>$s$</td>
<td>$\hat{p}$</td>
</tr>
<tr>
<td>Population</td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$p$</td>
</tr>
</tbody>
</table>

$\mu$ is pronounced “mu”
$\sigma$ is pronounced “sigma”

**Parameter**: a number that describes a population. Examples are $\mu$, $\sigma$, and $p$.

**Statistic**: a number that describes a sample. Examples are $\bar{x}$, $s$, $\hat{p}$, and $n$.

A normal curve visually describes a normal distribution. We draw a mathematical model normal curve to represent a normal population distribution. The curve is then used as an approximation to real life normal distributions and is accurate enough for practical purposes.

**Properties of a normal distribution:**
- Is symmetric (so mean = median = center).
- Has one major peak and no outliers.
- Mathematical model (shown to the right) has the x-axis as a horizontal asymptote.
- Area under normal curve equals 1.00. This corresponds to 100% of the data falling below the curve. For a normal distribution the area under a piece of the curve corresponds to the percentage of data along the x-axis under that curve. For example, if the area under a piece of the normal curve is 0.34 then the percentage of data under the curve would be 34%.

If $X$ represents a quantitative variable that is normally distributed then we may write $X \sim N(\mu, \sigma)$.

Example: $X \sim N(17, 3)$ means that the variable $X$ has a normal distribution with mean $= \mu = 17$ and standard deviation $= \sigma = 3$.

**Empirical Rule**: THE 68-95-99.7 RULE (Applies only to normal distributions!)
- About 68% of the data are within one standard deviation from the mean.
  In other words, about 68% of the data fall in the interval $\mu \pm 1\sigma$.
- About 95% of the data are within two standard deviations from the mean.
  In other words, about 95% of the data fall in the interval $\mu \pm 2\sigma$.
- About 99.7% of the data are within three standard deviations from the mean.
  In other words, about 99.7% of the data fall in the interval $\mu \pm 3\sigma$. 

$X$

$\mu - 3\sigma \quad \mu - 2\sigma \quad \mu - \sigma \quad \mu \quad \mu + \sigma \quad \mu + 2\sigma \quad \mu + 3\sigma$
To calculate the percentage of data under the normal curve between a given lower bound (LB) and upper bound (UB) use normalcdf on your calculator and enter in the following:

\[
\text{normalcdf} \ (\text{LB}, \text{UB}, \text{mean of the distribution}, \text{standard deviation of the distribution})
\]

If the mean and standard deviation are not entered then the calculator will assume that they are 0 and 1, respectively. Extreme lower bound (like negative infinity) is -1E99 on calculator. Extreme upper bound (like positive infinity) is 1E99 on calculator. To get E on the calculator press [2ND] COMMA. The comma button is above the 7 button.

\[\begin{align*}
\text{INFO IN PARAGRAPH:} & \quad \text{WANT TO FIND:} \\
\text{LB, UB, mean, standard deviation} & \quad \text{normalcdf} \\
& \quad \begin{cases}
\text{Percentage of data (in decimal form) between LB and UB.} \\
\text{Proportion of data between LB and UB.} \\
\text{Probability of choosing one data between LB and UB.} \\
\text{In notation: } P(\text{LB} < X < \text{UB}) \\
\text{Area under the curve between LB and UB.}
\end{cases}
\end{align*}\]

**Commuter times...**
1. Assume that commute times of students to school from home are normally distributed with a mean of 17 minutes and a standard deviation of 3 minutes.
   a. What percent of students take between 11 and 23 minutes to commute to school?
   b. What is the probability that a randomly chosen student commutes between 17 and 23 minutes?
   c. Pat Murphy takes about 20 to 23 minutes to get to school. What is the area under the curve between 20 to 23 minutes?
   d. What proportion of students take less than 14 minutes to get to school?

**IQ scores...**
2. Suppose that a data set of IQ scores are normally distributed with a mean of 110 and a standard deviation of 10. In symbols: \(X \sim N(110, 10)\).
   a. What percentage of IQ scores are between 100 and 120?
   b. What percentage of IQ scores are between 100 and 113?
   c. What percentage of IQ scores are lower than 95?
   d. What is the probability of picking an IQ score between 90 and 110?
   e. Calculate \(P(112 < X < 115)\).
   f. What is the probability of an IQ score being above 130?
   g. What proportion of IQ scores are greater than 110?
   h. What is the area under the normal curve between IQ scores of 103 and 123?

To calculate the unknown data \(X\), given the area to the left of \(X\) (in decimal form), use Invnorm on your calculator and enter in the following:

\[
\text{Invnorm} \ (\text{Area to the left, mean of the distribution, standard deviation of the distribution})
\]

If the mean and standard deviation are not entered then the calculator will assume that they are 0 and 1, respectively. Use InvNorm when you are asked to find an unknown number along the horizontal axis that borders a given area.
WANT TO FIND:  INFO IN PARAGRAPH:

\[ X \leftarrow \text{InvNorm} \]

- Percentage of data (in decimal form) to the left of unknown data X.
- Proportion of data to the left of unknown data X.
- Probability of choosing one data to the left of unknown data X.
- Area under the curve to the left of unknown data X.

Use \text{InvNorm} when you are asked to find an unknown number along the horizontal axis that borders a given area!

**Commute times continued...**

- Exactly 10% of students take longer than Ricardo. What is Ricardo’s commute time?
- How long does it take you to commute to school if your commute time is the 75th percentile of all other students?
- What commute times are in the lower 20% of all commute times?

**IQ scores continued...**

- What is the lowest IQ score you can have and still be in the top 15%?
- What IQ score do you have to have in order to be in the highest 1% of all IQ scores?
- What is the lowest IQ score that is above 60% of all the others?
- What is the value of \( Q_1 \) of all the IQ scores?

Standardized scores are called “z-scores.” Every data x can be standardized into a z-score called \( z \) (also \( z_x \)).

**Conversion formula:**

\[
\begin{align*}
  z &= \frac{x - \mu}{\sigma}
\end{align*}
\]

In general:

\[
\text{z-score} = \frac{\text{data value} - \text{mean of the distribution}}{\text{standard deviation of the distribution}}
\]

The z-score for an \( x \) data value is how many standard deviations that \( x \) value is away from the mean.

Z-scores can be thought of as falling on the **Standard Normal Curve**: the normal curve with mean = 0 and standard deviation = 1. In symbols: \( z \sim N(0,1) \). The area the data value borders from its normal curve is the same as the area its z-score borders on the Standard Normal Curve.

**IQ scores continued...**

- What is the z-score for an IQ score of 95?
- What is the z-score for an IQ score of 127?
- What IQ score corresponds to a z-score of \(-2.6\)?
- What are the top 5% of z-scores?
- Is the z-score for \( Q_1 \) negative, positive or zero?
- What is the z-score for an IQ score that is 3.1 standard deviations below the mean?