6.2 - *Continuous Probability Distributions*

Characteristics of a Continuous Probability Distribution

1. The **probability** that the continuous random variable X will assume a value between two possible values, X = a and X = b of the variable **is equal to the area** under the density curve between the values X = a and X = b.
2. The **total area** (or **probability**) under the density curve is equal to **1**.
3. The **probability** that the continuous random variable X will assume a value between **any two possible values** X = a and X = b of the variable is **between 0 and 1.**

Probability Statements Associated with a Continuous Probability Distribution

1. The probability of obtaining a specific value X = a of a continuous variable is zero. That is, $P(X = a) = 0$.
2. The probability that a continuous variable X assumes a value between a and b, written $P(a < X < b)$, and the probability that a continuous variable X assumes a value from a to b, written $P(a \leq X \leq b)$, are equal. That is, $P(a < X < b) = P(a \leq X \leq b)$.

6.3 - *The Normal Distribution*

- A normal distribution is a distribution that represents the values of a continuous variable.
- The importance of a normal distribution is that it serves as a *good model* in approximating many distributions of real-world phenomena.
- If a continuous variable is said to be approximately normal, then the normal distribution can be used to model the continuous variable.
- The graph of a normal distribution is called a normal curve which is a smooth symmetric bell-shaped curve with a single peak at the center.
- Examples of distributions that can be modeled or approximated by the normal distribution include IQ scores of individuals, adult weights, blood pressure of men and women, tire wear, the size of red blood cells, and the time required to get to work.

6.4 - *Properties of a Normal Distribution*

1. The normal curve is bell-shaped and has a single peak which is located at the center.
2. The mean, median, and mode all have the same numerical value and are located at the center of the distribution.
3. The normal curve is symmetric about the mean.
4. The data values in the normal distribution are clustered about the mean.
5. In theory, the normal curve extends infinitely in both directions, always getting closer to the horizontal axis but never touching it. As the normal curve extends out away from the mean and gets closer to the horizontal axis, the area in the tails of the normal curve is getting closer to zero.
6. The total area under the normal curve is 1 which can also be interpreted as probability. Thus, the area under the normal curve represents 100% of all the data values.

The Standard Normal Curve

- For each continuous variable that can be approximated by a normal distribution with a particular mean and standard deviation, there is a normal curve having the same mean and standard deviation which serves as a model of this variable. Consequently, there is no single normal curve, but rather many normal curves which are referred to as a family of curves.

- Within this family of normal curves, there is only one normal distribution or curve for each combination of mean and standard deviation and the number of possible normal curves is infinite.
- The proportion of area under any normal curve between the mean and the data value which is N standard deviations away from the mean is the same for all normal curves. Consequently, when working with normal distribution applications it is possible only to have to work with one normal curve since each member of the family of normal distributions exhibits the same characteristics.
- Although there are an infinite number of normal curves, there is one special member of the normal distribution family that can be used to serve as the normal distribution model for any application. This special normal distribution is referred to as the **standard normal distribution**.
- The standard normal distribution has a mean equal to zero, $\mu = 0$, and a standard deviation equal to one, $\sigma = 1$. To utilize this standard normal distribution as a model for any normal distribution, it becomes necessary to convert any normal distribution to the standard normal distribution. This is accomplished by converting the values of the continuous variable to z scores.
- The z score formula helps to convert any normal distribution to the standard normal distribution. The z score of a data value of x is equal to $\frac{x-\mu}{\sigma}$. Regardless of the values and units of any normal continuous variable, the z score formula can be used to transform any normal distribution to the standard normal distribution with a mean of 0 and a standard deviation of 1.
- Since any normal distribution can be converted to the standard normal distribution, we only need to refer to one statistical table to find the area under any normal curve rather than an infinite number of tables for all the possible normal curves. This table is called the Standard Normal Curve Area Table.

6.5 - *Using the Normal Curve Area Table*

Procedure to Determine the Proportion of Area Under a Normal Curve to the Left of a z Score in Table II

1. Determine which page of Table II to use.
2. Find the row corresponding to the z score under the column labeled z. This row is determined by looking for the integer value and the value of the first decimal place of the z score.
3. Locate the column corresponding to the z score. To determine this column, look for the value of the second decimal place of the z score by moving across the top row of the table until you find the column pertaining to the second decimal place.
4. The entry found in the table where the row and the column determined in steps 2 and 3 intersect represents the proportion of the area to the left of the z score.

Procedure to Find the Proportion of Area to the Left of a z Score

Proportion of area to the left of z = Entry in Table II corresponding to z.

Using TI83/84 calculator: [2nd] [DISTR] [2: normalcdf (] – [***EE99***] [**,**] [***z score***]) [ENTER]

Procedure to Find the Proportion of Area to the Right of a z Score

Proportion of area to the right of z = 1 – Proportion of area to the left of z.

Using TI83/84 calculator: [2nd] [DISTR] [2: normalcdf (] [***z score***] [**,**] [2nd] [***EE99***]) [ENTER]

Procedure to Find the Proportion of Area Between Two z Scores

Proportion of area between two z scores = Proportion of area to the left of larger z score – Proportion of area to the left of the smaller z score.

Using TI83/84 calculator: [2nd] [DISTR] [2: normalcdf (] [***smaller z score***] [**,**] [***larger z score***]) [ENTER]

Proportion of Area Associated to a Single z Score
Proportion of area for a particular z score equals zero.

Example 6.1 on pg. 304 Find the proportion of area under the normal curve:

a. To the left of z = – 0.53 (or Less than a z score of – 0.53)

Using TABLE II: Entry in table for z = – 0.53 is **0.2981** which represents the proportion of area to the **left** of z = – 0.53

Using your TI83/84 calculator:
[2nd] [DISTR] [2: normalcdf (] – [*EE99*] [,] [*z score*]) [ENTER]
normalcdf (– E99, – 0.53) = 0.2981

b. To the left of z = 2.56 (or Less than a z score of 2.56)

Using your TI83/84 calculator:
normalcdf (– E99, 2.56) = __________

Example 6.2 on pg. 306 Find the proportion of area under the normal curve:

a. To the right of z = – 1.37 (or greater than a z score of – 1.37)

Using TABLE II: Entry in table for z = – 1.37 is 0.0853 which represents the proportion of area to the left so
Proportion of area to the right of z-score = **1** – Proportion of **area** to the **left** of z
Proportion of area to the **right** of z = – 1.37 is **1 – 0.0853** = **0.9147**

Using your TI83/84 calculator:
[2nd] [DISTR] [2: normalcdf (] [*z score*] [,] [*EE99*]) [ENTER]
normalcdf (– 1.37, E99) = 0.9147

b. To the right of z = 0.67 (or greater than a z score of 0.67)

Using your TI83/84 calculator:
normalcdf (0.67, E99) = __________

Example 6.3 on pg. 308 Find the proportion of area under the normal curve:

a. Between z = 0 and z = 1.5

Using TABLE II:
Entry in table for z = 0 is 0.5000 which represents the proportion of area to the left of z = 0
Entry in table for z = 1.5 is 0.9332 which represents the proportion of area to the left of z = 1.5
Proportion between two z scores = Proportion of **area** to the left of **larger z score** – of **area** to the left of the **smaller z score**
Proportion of area **between** z = 0 and z = 1.5 is **0.9332 – 0.5000** = **0.4332**

Using your TI83/84 calculator:
[2nd] [DISTR] [2: normalcdf (] [*smaller z score*] [,] [*larger z score*]) [ENTER]
normalcdf (0, 1.5) = 0.4332

b. Between z = – 1.96 and z = 1.96

Using your TI83/84 calculator:
normalcdf (– 1.96, 1.96) = **__________**

c. Between z = – 1.25 and z = 1.0

Using your TI83/84 calculator:
normalcdf (– 1.25, 1.0) = __________

Example 6.5 on pg. 313 For a normal distribution, find the z score(s) that cut(s) off:

a. The lowest 20% of area. (or cuts off the bottom 20 % or 0.20 of the z scores)

Using your TI83/84 calculator:
20% or 0.20 proportion of area to the left
invNorm (0.20) = – 0.84

b. The highest of 10% of area (or cuts off the top 10 % or 0.10 of the z scores)

Using your TI83/84 calculator:
90% or 0.90 proportion of area to the left
invNorm (0.90) = 1.28

c. The middle 90% of the area (or of the z scores)

There will be **two** z-scores as your answer (one positive and one negative)

Using your TI83/84 calculator:

5% or 0.05 proportion of area to the left of lower z-score
invNorm (0.05) = – 1.64
and
95% or 0.95 proportion of area to the left of the higher z-score
invNorm (0.95) = 1.64

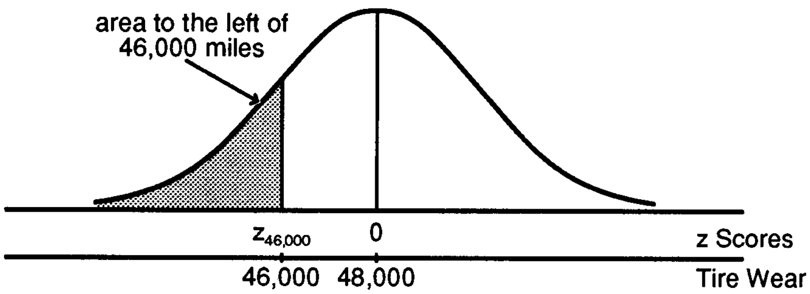
***There will always be two z-scores which are the same number but opposite signs when you cut off the "middle".**

6.6 - *Applications of the Normal Distribution*

Example 6.6 on pg. 316 Assume the distribution of tire wear is approximately normal with: $\mu = 48{,}000; \sigma = 2{,}000$. Find the **proportion** of tires which would be expected to wear:

a. Less than 46,000 miles
b. Greater than 49,000 miles
c. Between 47,000 and 51,000 miles

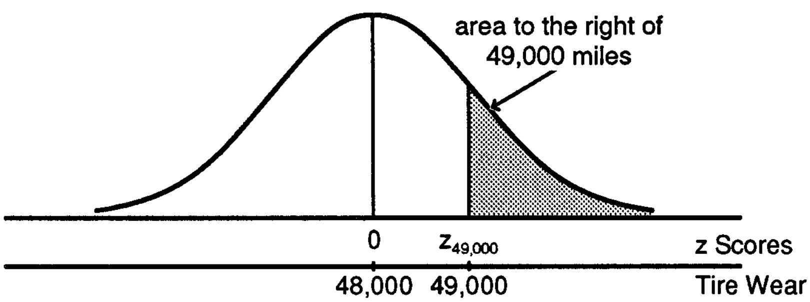
a. Less than 46,000 miles



Using your TI83/84 calculator:
[2nd] [DISTR] [2: normalcdf (] [lowest data value] [,] [highest data value] [,] [μ] [,] [σ] [ENTER]
normalcdf (– E99, 46000, 48000, 2000) = __________

ANS: The the proportion of tires that is expected to wear less than 46,000 miles is __________

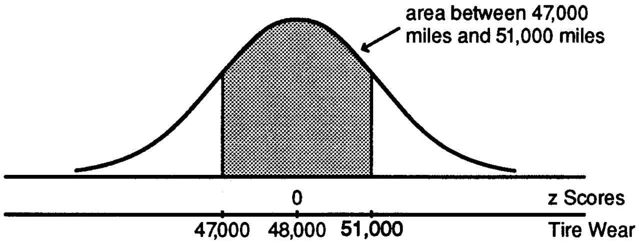
b. Greater than 49,000 miles



Using your TI83/84 calculator:
normalcdf (49000, E99, 48000, 2000) = __________

ANS: The proportion of tires that is expected to wear more than 49,000 is __________

c. Between 47,000 and 51,000 miles



Using your TI83/84 calculator:
normalcdf (47000, 51000, 48000, 2000) = __________

ANS: The proportion of tires that is expected to wear between 47,000 and 51,000 is __________

Note: If asked for the **percent** of tires which would be expected to wear 47,000 and 51,000 the answer would be **62.47%**

Review ***Example 6.7 on pg. 317*** (parts **a** and **b**) in the text.

Example 6.8 on pg. 319 One thousand students took a standardized psychology exam. The results were approximately normal with $\mu = 83$ and $\sigma = 8$. Find the ***number*** of students who scored:

a. Less than 67

Using your TI83/84 calculator:
normalcdf (– E 99, 67, 83, 8) = 0.0228

To find the **number** of students and not the proportion of area we must multiply the proportion with the total number of students that took the test.
(0.0228) (1000) = 22.8 or **23** students scored less than 67

b. Greater than 87
normalcdf (__________ , __________ , __________ , __________) = __________

6.7 - *Percentiles and Applications of Percentiles*

- **Percentiles** are widely used to describe the position of a data value within a distribution.

The **percentile rank** of a data value X within a normal distribution is equal to the percent of area under the normal curve to the left of the data value X. **Percentile ranks are always expressed as a whole number.**

Procedure to Find Data Value (or Raw Score) that Cuts off Bottom p% Within a Normal Distribution

1. Change the percentage p to a proportion.
2. Using Table II, find the z score that has the area closest to this proportion. If there is a tie, choose the z score with the larger absolute value.
3. Substitute this z score into the raw score formula $X = \mu + (z)\sigma$ to determine the data value that corresponds to this z score.

Procedure to Find Data Value (or Raw Score) that Cuts off Top q% Within a Normal Distribution

1. Calculate the percentage of data values below the raw score by subtracting q from 100: (100 – q).
2. Change this percentage to a proportion.
3. Using Table II, find the z score that has the area closest to this proportion. If there is a tie, choose the z score with the larger absolute value.
4. Substitute this z score into the raw score formula $X = \mu + (z)\sigma$ to determine the data value that corresponds to this z score.

For percentile rank – always use proportion of **area to the left** of the **data value** and **always** round to the nearest **whole number**.

Using your TI83/84 calculator for <u>percentile rank</u>:

2^{nd} DISTR 2: normalcdf (– *EE99* , data value , μ , σ) ENTER

Example 6.9 on pg. 320 Matthew earned a grade of 87 on his history exam. If the grade in his class were normally distributed with $\mu = 80$ and $\sigma = 7$, find Matthew's **percentile rank** on this exam.

Using your TI83/84 calculator for <u>percentile rank</u>:
2^{nd} DISTR 2: normalcdf (– ***EE99*** , data value , μ , σ) ENTER

normalcdf (– E 99, 87, 80, 7) = 0.8413, now change this to a percent.

(proportion of area to the left) (100) = 84.13% , now round to the nearest whole number.

Percentile Rank = 84

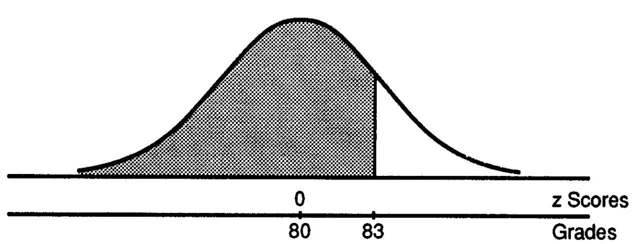
Notation for this is: $87 = P_{\mathbf{84}}$

Example 6.10 on pg. 321 Miss Brooke took two exams. Her grades and the class results are as follows:

<u>Exam</u>	Her <u>Grade</u>	Class <u>mean (μ)</u>	Class Standard <u>Deviation (σ)</u>
English	83	80	6
Math	77	74	5

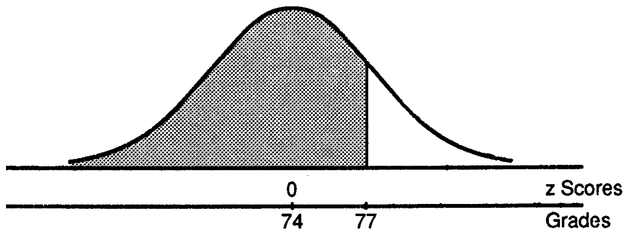
Use percentile ranks to determine on which test(s) Miss Brooke did better relative to her class

For English:



normalcdf (– E 99, 83, 80, 6) = __________= __________% = percentile rank is __________
The **percentile rank** of her English grade of 83 is **69**, expressed as $P_{69} = 83$

For Math:



normalcdf (__________ , __________ , __________ , __________) = __________
As a percent: __________%
The percentile rank is __________
The **percentile rank** of her Math grade of 77 is __________, expressed as P =

Answer/Conclusion: The percentile rank of her math grade is greater than the percentile rank of her English grade so relative to her class, Miss Brooke did **better on her math exam**.

Example 6.11 on pg. 322 One of the entrance requirements at a local college is a percentile rank of at least 45 on the verbal SAT exam. Bill scored 470 on his verbal SAT exam. Does he meet the minimum requirement of this college, if the verbal SAT distribution is normally distributed with $\mu = 500$ and $\sigma = 100$?

6.8 - *Probability Applications*

Probability of a data value X falling between two data values: X_1 and X_2.
The probability that a data value X selected at random from a normal distribution will fall between the two data values X_1 and X_2 is equal to the proportion of area between X_1 and X_2. This can be expressed as: $P(X_1 \le X \le X_2) =$ **Area under the normal curve between** X_1 and X_2

Example 6.17 on pg. 332 Assume the distribution of the playing careers of major league baseball players can be approximated by a normal distribution with a mean of 8 years and a standard deviation of 4 years. Find the probability that a player selected at random will have a career that will last:

a. Less than 6 years
Using your TI83/84 calculator:
[2nd] [DISTR] [2: normalcdf (] [lowest data value] [,] [highest data value] [,] [μ] [,] [σ] [ENTER]
normalcdf (__________ , __________ , __________ , __________) = __________

b. Longer than 14 years
Using your TI83/84 calculator:
normalcdf (__________ , __________ , __________ , __________) = __________

c. Between 4 and 10 years
Using your TI83/84 calculator:
normalcdf (__________ , __________ , __________ , __________) = __________

6.9 - *The Normal Approximation to the Binomial Distribution*

Definitions

n = number of trials

p = the probability of success in one trial

$1 - p$ = the probability of failure in one trial

μ_s = *mu sub s* = $\mu_s = np$

σ_s = *sigma sub s* = $\sqrt{np(1-p)}$

Required calculations for the Normal Approximation to the Binomial Distribution

Conditions of Normal Approximation of a Binomial Distribution

You can use the approximation by a normal distribution, with a mean of μ_s and a σ_s when **both**:

np **and** $n(1-p)$ **> 5**

Using your TI83/84 calculator:

[2nd] [DISTR] [2: normalcdf (] [lower value] [,] [higher value] [,] [μ] [,] [σ] [ENTER]

normalcdf (lower value, upper value, μ_s, σ_s)

* **Must use the correction for continuity on the lower value and/or higher value (see table on pg 11 of handout)**

Example 6.20 on p.341 Consider the binomial experiment of tossing a fair coin 100 times. Use the normal approximation to calculate the probability of getting:

a. At least 55 heads
b. Between 40 and 60 heads
c. Exactly 54 heads
d. At most 45 heads

$n = 100$ $p = ½$ $1 - p = 1 - ½ = ½$

np and $n(1-p)$ are each > 5 so we can use normal

Find:

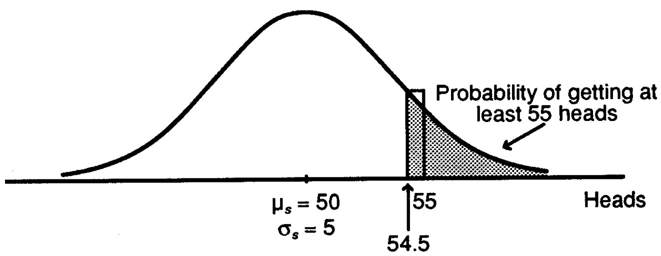
$\mu_s = np = 100 \times (.5) = 50$

and

$\sigma_s = \sqrt{np(1-p)} = \sqrt{50(.5)} = \sqrt{25} = 5$

a. Probability of getting at least 55 heads

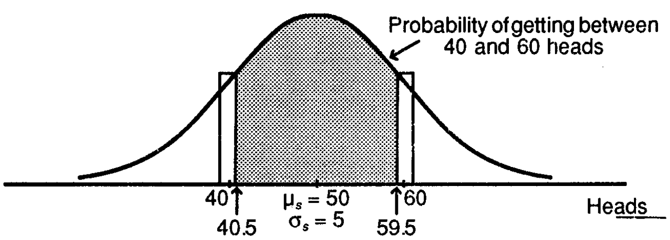
Use 54.5, which is the lower boundary of 55 and the **correction for continuity**



We are looking for the area under the curve that is to the right of 54.5.
$\mu_s = 50$ and $\sigma_s = 5$
normalcdf (lower value, upper value, μ_s, σ_s)
normalcdf (54.5, E99, 50, 5) = **0.1841**
So the probability of getting at least 55 heads is approximately **0.1841**

b. Probability of getting between 40 and 60 heads (not inclusive)

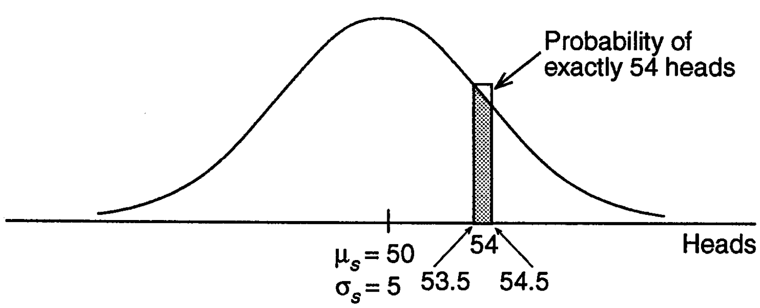
Use the upper boundary of 40, which is 40.5 and the lower boundary of 60, which is 59.5



normalcdf (40.5, 59.5, 50, 5) = **0.9426**

c. Probability of getting exactly 54 heads

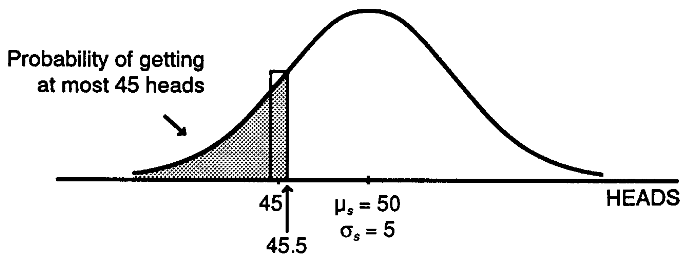
Use the lower boundary of 54, which is 53.5 and the upper boundary of 54, which is 54.5



normalcdf (53.5, 54.5, 50, 5) = **0.0579**

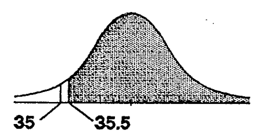
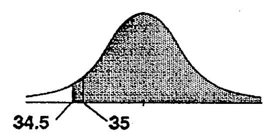
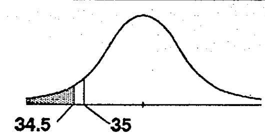
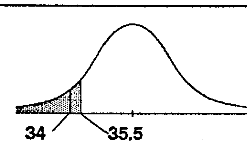
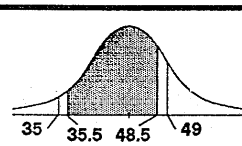
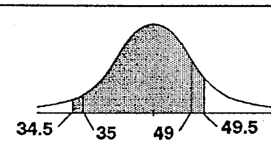
d. Probability of getting at most 45 heads

This is the same as saying getting ≤ 45 heads
Use 45.5, which is the upper boundary of 45



normalcdf (– E99, 45.5, 50, 5) = **0.1841**

Review ***Examples 6.21 & 6.22 on pg. 344*** in the text.

KEY WORD PHRASE	SYMBOLIZED STATEMENT S = Number of Successes	REAL BOUNDARIES	Diagram of area Indicated by KEY WORD Phrase
More than 35 Greater than 35	P(S > 35)	35 + .5 = 35.5	
At least 35 Greater than or equal to 35 35 and more 35 or more	P(S ≥ 35)	35 - .5 = 34.5	
Less than 35 Fewer than 35	P(S < 35)	35 - .5 = 34.5	
At most 35 Less than or equal to 35 35 and less 35 or less	P(S ≤ 35)	35 + .5 = 35.5	
Between 35 and 49	P(35 < S < 49)	Lower: 35 + .5 = 35.5 Upper: 49 - .5 = 48.5	
Between 35 and 49 Inclusive	P(35 ≤ S ≤ 49)	Lower: 35 - .5 = 34.5 Upper: 49 + .5 = 49.5	
Exactly 35	P(S = 35)	Lower: 35 - .5 = 34.5 Upper: 35 + .5 = 35.5