Section 3.2: Measures of Spread

The mean and median are good statistics to employ when describing the center of a collection of data. However, there is more to a collection of data than just the center! Recall our example of average test scores in two different classes.

Example 1 *The average score on Test 1 in MATH 8000 at the University of Nowhere is 75. Of the 100 students in the class, half scored a 50 and the other half scored 100.*

Example 2 *The average score on Test 1 in MATH 8000 at the University of Nowhere is 75. Each of the 100 students in the class scored a 75.*

Problem 3 *What is the median score in each class?*

Knowing the mean and median is a good start to understanding data. But it is not enough. We must also understand how data varies. The same unit of measurement may not always have the same value or meaning.

Example 4 *Consider two teams of five cross country runners. Both teams average a 5 minute mile. Who wins a mile race? Consider the mile times (in minutes) for each team member given in the following table.*

Team 1	*Alice*	*Bob*	*Chris*	*David*	*Emily*
mile time	*5*	*5*	*5*	*5*	*5*
Team 2	*Frank*	*Greg*	*Hannah*	*Ian*	*Jenny*
mile time	*5*	*6*	*5*	*5*	*4*

While both teams average a 5 minute mile, Jenny wins the race for her team. Knowing the center of a collection of data is important but there is also the need to understand how the data varies.

1 Range

The simplest measure of the **spread** (or **dispersion**) of data is the range.

Definition 5 *For a given set of data, the* ***range*** *is the (positive or occasionally 0) difference between the largest and smallest values in a quantitative data set.*

Example 6 *The range of mile times for team 1 is* $5 - 5 = 0$. *The range of mile times for team 2 is* $6 - 4 = 2$.

Problem 7 *What is the range of salaries for the Chicago Bull's '97-'98 roster?*

	Player	Salary
1	Michael Jordan	$33,140,000
2	Ron Harper	$4,560,000
3	Toni Kukoc	$4,560,000
4	Dennis Rodman	$4,500,000
5	Luc Longley	$3,184,900
6	Scottie Pippen	$2,775,000
7	Bill Wennington	$1,800,000
8	Scott Burrell	$1,430,000
9	Randy Brown	$1,260,000
10	Robert Parish	$1,150,000
11	Jason Caffey	$850,920
12	Steve Kerr	$750,000
13	Keith Booth	$597,600
14	Jud Buechler	$500,000
15	Joe Kleine	$272,250
	Average	$4,088,711
	Median	$1,430,000

Chicago Bulls Salaries 1997-1998 Season

Problem 8 *Is "range" a resistant function?*

2 Variance and Standard Deviation

The most important and commonly used measure of spread is the standard deviation.

Definition 9 *Standard deviation measures the spread of the data from the mean. This can be seen at the heart of the formula for standard deviation. We denote a sample standard deviation by s and a population standard deviation by σ. There is a subtle difference in the two formulae.*

$$\begin{aligned} s &= \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \\ \sigma &= \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \end{aligned}$$

Definition 10 *Variance for samples and population is s^2 and σ^2.*

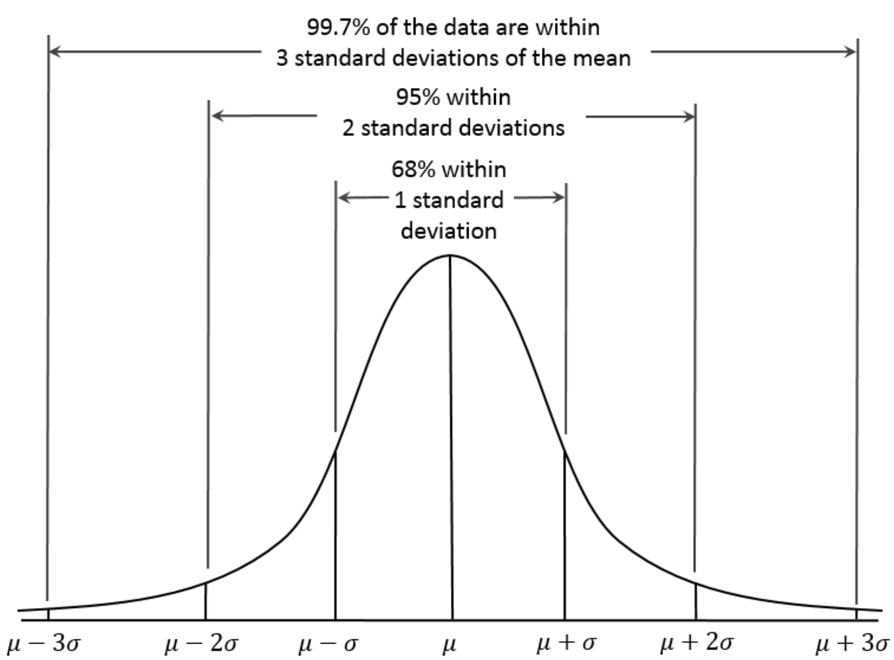
Know how to compute variance and standard deviation on the TI 83/84. For the player salaries on the 1997-1998 Chicago Bulls the sample standard deviation is $8,182,474.38 and the population standard deviation is.$7,905,021.27.

Example 11 *What is the sample variance for Bull's salaries? Variance is always standard deviation squared. So, sample variance for Bull's salaries is* $s^2 = 8182474.38^2 = 6.6953 \times 10^{13}$.

3 The Empirical Rule (68-95-99.7 Rule)

For data distributions that have a **bell-shape distribution** (**normal curve**), the mean and standard deviation tell us a lot about the spread of data from the center.

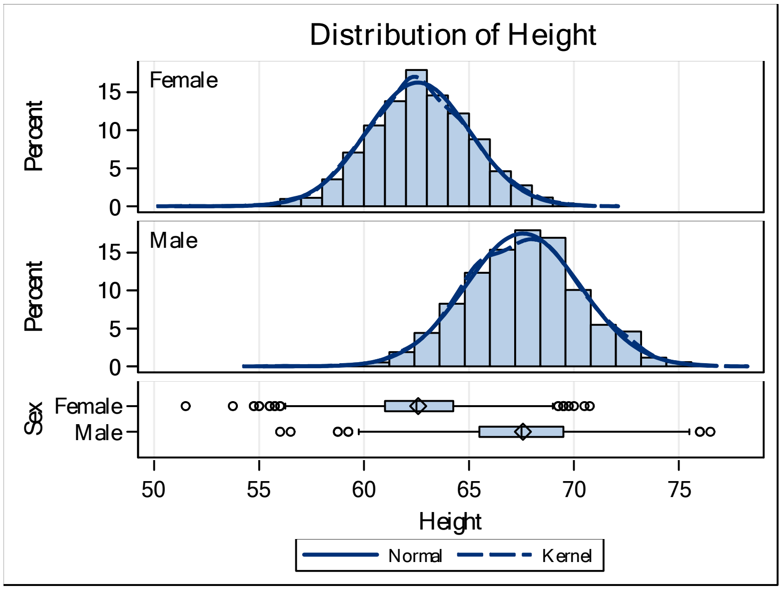
Theorem 12 *The Empirical Rule (68-95-99.7 Rule) states that every normal distribution 68% of the data falls within one standard deviation of the mean, 95% of the data falls within two standard deviations of the mean and 99.7% of the data (almost all) falls within three standard deviations of the mean.*



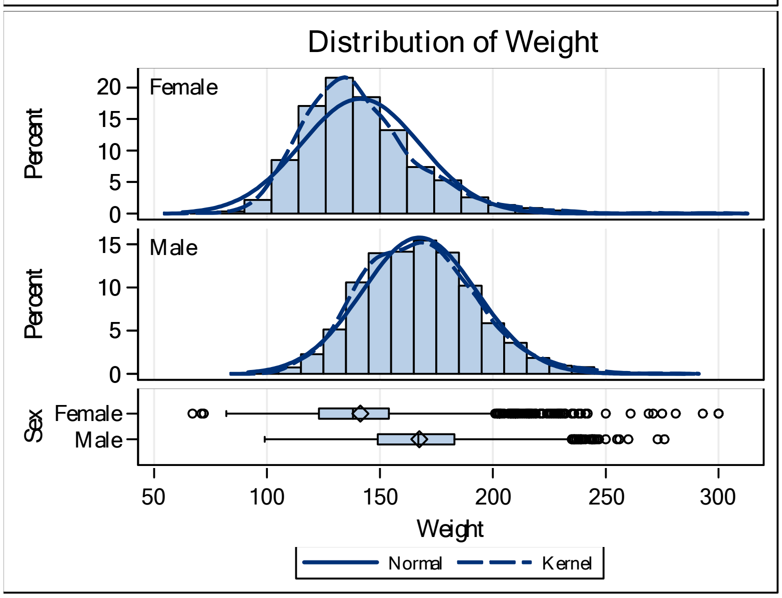
https://en.wikipedia.org/wiki/Normal_distribution

Remark 13 *We frequently denote a normal distribution by* $N(\mu, \sigma)$.

Example 14 *Heights and weights of men and women follow a normal distribution.*

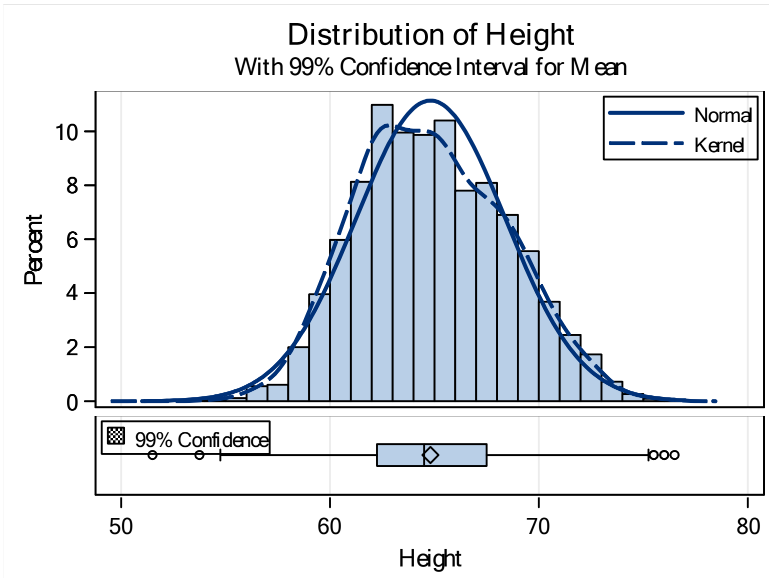


Distribution of Height
Female
Male
Percent
Sex
Height
Normal
Kernel



Distribution of Weight
Female
Male
Percent
Sex
Weight
Normal
Kernel

Remark 15 *Note that we needed to seperate the men and women into different groups for a normal distribution to appear.*



Problem 16 *Heights of men follow a normal distribution with an average of 69" and a standard deviation of 2.8" (I'm rounding to make the arithmetic easier).*

This indicates that the central 68% of men are from   to   tall.

What percentage of men are between 60.6" and 77.4" tall?

What percentage of men are between 69" and 71.8" tall?

Problem 17 *Consider a class whose test results have a distribution of N(75,6).*

What grades comprise the central 68% of the students?

What percentage of grades are between 75 and 81?

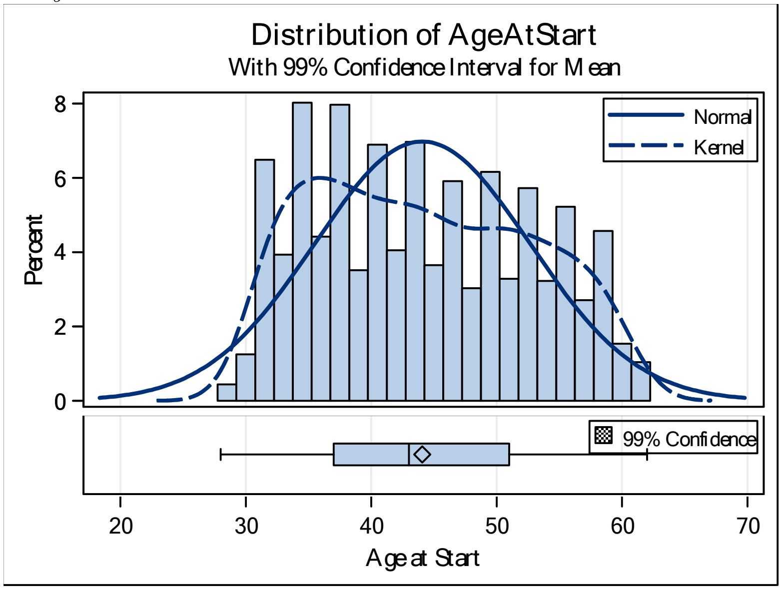
What percentage of grades are between 81 and 87?

What percentage of grades are below 57?

If 2000 students took this test, how many students earned a grade between 81 and 87?

4 Chebyshev's Inequality

Example 18 *The distribution of age of patient at the beginning of the health study is not normal.*



Not every distribution of data is bell-shaped. In such cases we can use Chebyshev's inequality to approximate the percentage of data that falls within a certain number of standard deviations of the mean.

Theorem 19 *For every distribution of data, at least* $1 - \frac{1}{k^2}$ *percent of the data falls within* k *standard deviations of the mean.*

Typically, we only use Chebyshev's inequality for $k = 2$ or $k = 3$.

At least $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = .75 = 75\%$ of the data falls within 2 standard deviations of the mean.

At least $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = .889 = 88.9\%$ of the data falls within 3 standard deviations of the mean.

Problem 20 *Consider a class whose test results have an average of 78 and a standard deviation of 10. This distribution is not bell-shaped.*

At least what percentage of students earned a grade between 58 and 98?

Problem 21 *Par on the Captain Kidd's Challenge in Hilton Head, SC is 56 strokes with a standard deviation of 12. At least 88.9% of the scores fall into what interval?*

5 Exercises

1. We have computed the mean, median and standard deviation for the 1997-1998 Chicago Bulls salaries. Suppose that every player receives a $1,000,000 raise? Find the values for min, max, mean, median and standard deviation for the post raise salaries. How have these values changed?

2. We have computed the mean, median and standard deviation for the 1997-1998 Chicago Bulls salaries. Suppose that every player receives a 10% raise? Find the values for min, max, mean, median and standard deviation for the post raise salaries. How have these values changed?

3. Let's play a game! Every student gets to play this game once. I have two boxes up front on my gaming table. A single play of this game consists of a student selecting a box and then randomly selecting a ticket from the box. The student then receives the value of that ticket.

   Box A

$0	$4	$500	$997
$1	$5	$994	$998
$2	$6	$995	$999
$3	$500	$996	$1000

   Box B

$0	$500	$500	$500
$500	$500	$500	$500
$500	$500	$500	$500
$500	$500	$500	$1000

   Which box will you pick (There is no wrong answer)? Compute mean, median and standard deviation and use some or all of these values to back up your answer.

4. Navidi/Monk Section 3.2: 9-12, 17-22, 27-34, 36, 41-48