Chapter 18 – Measures of Center

- Data that can be any numerical value are called <u>continuous</u>. These are usually things that are measured, such as height, length, time, speed, etc.

- Data that can only be integer values are called <u>discrete</u>. These are usually things that are counted, such as apples, owls, cars, books, etc.

- There are three statistics that are used to measure the center of a data set: mean, median, and mode.

  mean – the average value
  median – the middle value in an ordered data set
  mode – the data value that occurs most often

  Example: The data below represent the heights, in inches, of ten high school basketball players:

  65 66 66 67 67 68 68 68 70 75

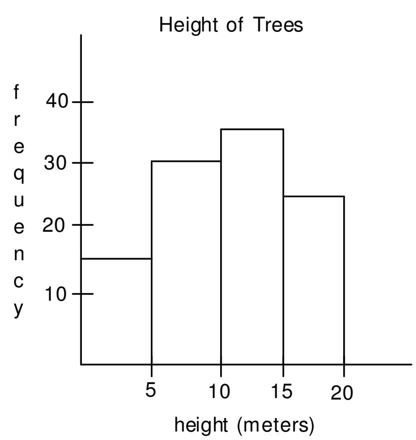
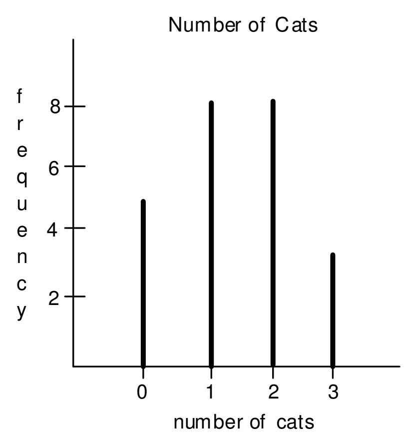
  Find the mean, median, and mode.

  mean = 68 median = 67.5 mode = 68

- An extreme value in a data set is called an <u>outlier</u>. Sometimes outliers are excluded before the data are analyzed.

- Some examples of displays of data:

<table>
<tr><th>Stem and
Leaf Plot</th><th>Histogram
(for continuous data)</th><th>Column Graph
(for discrete data)</th></tr>
<tr><td>stem | leaf
2 | 3 4 6 6
3 | 0 1 5
4 | 2 4 4 7 9

where 2|3 means 2.3</td><td>
</td><td>
</td></tr>
</table>

Chapter 18 – Measures of Dispersion

- Another way to analyze data is using measures of dispersion: range and interquartile range. These show how widely the data vary.

- The range is difference between the highest and lowest data values.

- The median divides the data set into two halves. The median of the lower half is called the lower quartile (Q1), and the median of the upper half is called the upper quartile (Q3).

- The interquartile range (IQR) is the difference between the upper and lower quartiles, Q3 – Q1.

- The data set is divided into quarters by the lower quartile (Q1), the median (Q2), and the upper quartile (Q3). 25% of the data are less than Q1, 50% are less than Q2, and 75% are less than Q3. These are also called percentiles.

- The upper and lower quartiles along with the median and the minimum and maximum values form the five number summary of the data set.

- A box and whisker plot is a type of graph that shows the dispersion of data, including the five number summary.

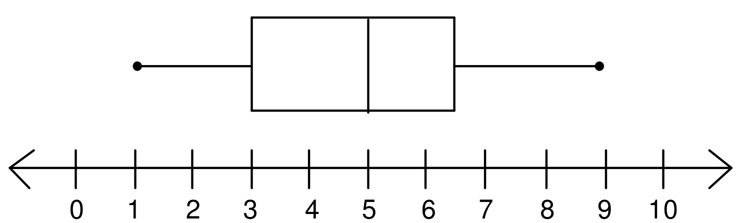
Example: For the data set below, find the five number summary, draw a box and whisker plot, and state the range and the interquartile range.

4 5 9 5 1 7 8 7 3 5 6 3 4 3 2 5

ordered data (n = 16)

1 2 3 3 3 4 4 5 5 5 5 6 7 7 8 9

min = 1
Q1 = 3
median = 5
Q1 = 6.5
max = 9



The range is 9 – 1 = 8. The interquartile range is 6.5 – 3 = 3.5.

Chapter 18 – Cumulative Frequency

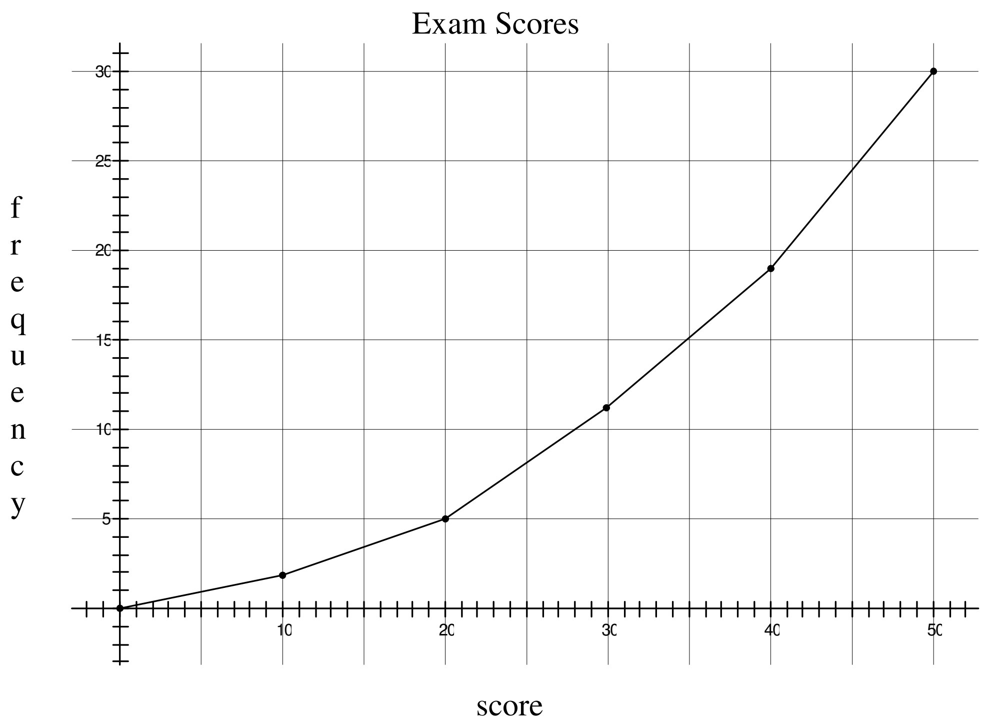
- A cumulative frequency distribution table shows the total frequency of data points, up to and including a particular value (or range of values).

Example:

Exam Score	Frequency	Cumulative Frequency
$0 \le x < 10$	2	2
$10 \le x < 20$	3	5
$20 \le x < 30$	6	11
$30 \le x < 40$	8	19
$40 \le x < 50$	11	30

- Cumulative data can also be represented using a cumulative frequency polygon graph. The graph can be used to estimate the median and to find other properties of the data.

Example:



The median is about 35 points.
If passing is a score of 30 points or more, about 19 students passed.
About 6 students earned an A (90% or more).
About 5 students earned a B (between 80% and 90%).
The 80th percentile is a score of about 44 points.

Chapter 18 – Variance and Standard Deviation

- Since range and interquartile range use only two data points, they are not very informative. So there are two other more important measures of dispersion that use all the data values: variance and standard deviation.

- The deviation is the difference between the data value (x) and the mean (μ). Every value in the data set has a deviation.

- Because some deviations are positive and some are negative, we square them and then find the average of the squared deviations. This is called the variance and is denoted σ^2.

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n}$$

- The standard deviation, σ, is the square root of the variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n}}$$

- Note that in some contexts $\overline{x}$ is used for the mean and s is used for the standard deviation.

- You can use your graphing calculator to find standard deviation, but be careful if you're doing an IB problem! For IB problems, you should always use the "1-var stats" function and chose the value labeled σ. Do not use the value labeled s or the "stdDev" function!

- Standard deviation is very important for data that are normally distributed, which we will study later.

Chapter 19 – Introduction to Probability

- Probability is the likelihood that something will happen. Probability can be measured numerically.

- A random experiment is an experiment in which there is no way to determine the outcome beforehand. For example, a dice game.

- A trial is an action in a random experiment. For example, rolling the dice.

- An outcome is a possible result of a trial. For example, rolling 2-4.

- An event is a set of possible outcomes. For example, the total of the two dice is six (the outcomes in this event are 1-5, 2-4, 3-3, 4-2, 5-1).

- The sample space is the set of all possible outcomes of a random experiment. In the dice game experiment, the sample space contains 36 outcomes, as shown in the grid on page 357.

- For events that are equally likely, the probability is the number of outcomes in the event divided by the total number of outcomes in the sample space.

  Example: $P(\text{the total of the two dice is six}) = \frac{5}{36} = 0.13\overline{8} = 13.8\ \%$

- An event that is certain to occur has a probability of one. An event that cannot occur has a probability of zero. Probability is always a number between zero and one, inclusive.

  Examples: $P(\text{the total of the two dice is at least one}) = \frac{36}{36} = 1$

  $P(\text{the total of the two dice is exactly one}) = \frac{0}{36} = 0$

- It is also sometimes helpful to illustrate the sample space, and there are a several ways to do this: a list, a grid, a tree diagram, and a Venn diagram. We will learn more about how to use these methods later in the chapter.

Chapter 19 – Properties of Probability

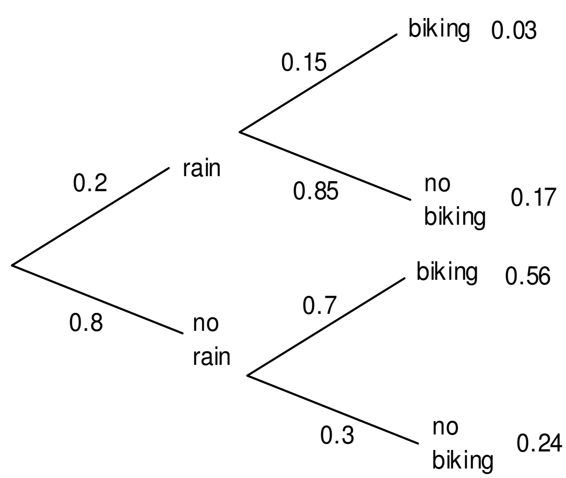
- Events are <u>complementary</u> if their probabilities add up to one. This means that one of the events is certain to happen.

  Example: The probability of rain on Tuesday is 0.2. What is the probability that it does not rain on Tuesday?

  These events are complementary, so 0.8.

- Tree diagrams are a useful tool for solving probability problems. When drawing a tree diagram, follow these guidelines:
  - Always draw tree diagrams horizontally.
  - Draw one set of branches for each action in the experiment.
  - Label the events at the end of the branches, and label the probabilities on the branches.
  - Multiply out on each branch to get the probability of each outcome.
  - The probabilities on each branch must always total to one <u>and</u> the final probabilities must always total to one.

  Example: There is a 20% chance of rain tomorrow. If it is raining, there is a 15% chance I will ride my bike after school. If it is not raining, there is a 70% chance I will go biking. Find the probability that I ride my bike after school tomorrow.



P(I go biking tomorrow) = 0.03 + 0.56 = 0.59

- Events that do not affect each other are called independent. For example, drawing two cards from a deck with replacement.

- Events that do affect each other are called conditional. For example, drawing two cards from a deck without replacement.

- The symbol $A \cap B$ means the intersection of *A* with *B*. It is equivalent to *A* and *B*.

- The multiplication law for probability says $P(A \cap B) = P(A) \cdot P(B/A)$ where $P(B/A)$ means the probability of *B* happening given *A* has happened.

- For independent events the multiplication law simplifies to $P(A \cap B) = P(A) \cdot P(B)$ if and only if the events are independent. This is because $P(B/A) = P(B)$ for independent events.

- The addition law for probability says $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- For mutually exclusive events the addition law simplifies to $P(A \cup B) = P(A) + P(B)$ if and only if the events are mutually exclusive. This is because $P(A \cap B) = 0$ for mutually exclusive events.

Example: Suppose $P(A) = 0.3$, $P(B) = 0.5$. Find $P(A \cup B)$ if:

a) the events are mutually exclusive
b) the events are independent

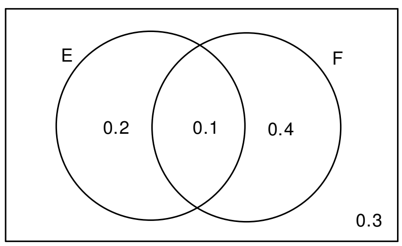
a) $P(A \cup B) = P(A) + P(B) = 0.3 + 0.5 = 0.8$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.3 + 0.5 - 0.3(0.5) = 0.65$

Example: Suppose $P(C)=0.2$, $P(D)=0.7$, and $P(C\cup D)=0.8$. Find $P(C\cap D)$.

$$P(C\cup D)=P(C)+P(D)-P(C\cap D)$$
$$0.8=0.2+0.7-x$$
$$0.8=0.9-x$$
$$x=0.1$$

Example: Event E and event F are shown in the Venn diagram below. Are these events independent? Explain.



$P(E)=0.3$ $\qquad$ $P(E)\cdot P(F)=0.15$

$P(F)=0.5$ $\qquad$ $P(E\cap F)=0.1$

Since $P(E\cap F)\neq P(E)\cdot P(F)$, the events are not independent.

Chapter 29 – The Normal Distribution

- The most important distribution for a continuous random variable is the normal distribution. It is given by a function of the form
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
defined for any real number *x*. This equation represents a family of functions that depend on μ (mean) and σ (standard deviation).

- The probabilities for a normal distribution are given by the area under the curve, and they are found using definite integrals:
$$P(a\le x\le b)=\int_a^b f(x)$$

- Note that because *X* is continuous, $P(X=x)$ is zero. Therefore it doesn't matter whether you use $\le$ or < since $P(X\le x)=P(X<x)$.

- Characteristics of the normal curve:
    - bell-shaped
    - symmetric about μ
    - inflection points at $\mu+\sigma$ and $\mu-\sigma$
    - asymptotic to the *x*-axis as *x* approaches infinity and negative infinity
    - the total area under the curve is one
    - the height and width of the curve depend on σ
    - 99.95% of the values are within ± 3.5 standard deviations of the mean
    - percentages of values are distributed as shown in the diagram on page 730

- The standard normal distribution has $\mu=0$ and $\sigma=1$. It is denoted by *Z* and is sometimes called the *z*-distribution.

- Probabilities for any normal distribution can be found using the graphing calculator. The input is normalcdf(min, max, μ, σ).

  Example: *X* is normally distributed with mean 70 and standard deviation 4. Find:
  a) $P(60<x<72)$ b) $P(x>76)$

  a) $P(60<x<72)=\text{normalcdf}(60,\ 72,\ 70,\ 4)\approx 0.685$
  b) $P(x>76)=\text{normalcdf}(76,\ 999,\ 70,\ 4)\approx 0.0668$

Chapter 29 – Inverse of the Normal Distribution

- When we are given the probability (or percentage of values) for a normally distributed random variable, we can find the particular value that gives that probability using the inverse of the normal distribution.

- On the GDC the inputs are invNorm(p, μ, σ). This gives the value of k for $P(X<k)=p$.

  Example: X is normally distributed with mean 32 and standard deviation 3. Find the value of k so that $P(X<k)=0.37$.

$$k=\text{invNorm}(0.37,32,3)\approx 31.0$$

- Note that the GDC only gives values for probabilities less than k. If you need to find the value of k for $P(X>k)=p$ you must use the complement, $P(X<k)=1-p$.

  Example: Scores on a physics exam are normally distributed $\mu=76$ and $\sigma=25$. The teacher decides to award an A to the top 7% of students. Find the minimum score required to earn an A.

$$P(X>k)=.07 \quad \Rightarrow \quad P(X<k)=.93$$

$$k=\text{invNorm}(0.93,76,25)\approx 82.9$$

The minimum score required to earn an A is about 83 points.

- The inverse of the normal distribution can also be used in problems where we know probabilities but μ and σ are unknown. To do this, we must first convert the x-values into z-values using the formula $z=\frac{x-\mu}{\sigma}$. This is known as standardizing.

  Example: X is normally distributed with $\sigma=22$. Given that $P(X<40)=.12$, find μ.

Since μ is unknown, we must first standardize the x-value, which in this problem is 40:

$$z=\frac{x-\mu}{\sigma}=\frac{40-\mu}{22}$$

Now $P(X<40)=.12$ becomes $P\left(Z<\frac{40-\mu}{22}\right)=.12$, and we can use the inverse normal button on the GDC:

$$\frac{40-\mu}{22}=\text{invNorm}(.12,\ 0,\ 1)$$

$$\frac{40-\mu}{22}=-1.17 \quad \Rightarrow \quad 40-\mu=-25.8 \quad \Rightarrow \quad \mu=65.8$$

Chapter 29 – The Binomial Distribution

- The binomial probability distribution applies to a random variable with the following characteristics:
  - The probability distribution is discrete.
  - There are a fixed number of trials.
  - There are exactly two outcomes, usually called "success" and "failure."
  - The probability of success is the same for all trials, that is, the trials are independent.

  The hardest thing about the binomial probability distribution is determining when to use it, so it is important to understand these characteristics!

- The symbol $X \sim B(n,\ p)$ means that *X* is a binomial distribution where *n* is the number of trials and *p* is the probability of success.

- The name binomial is used here because the formula used to calculate these probabilities is essentially the same as the binomial expansion formula. For $X \sim B(n,\ p)$:

$$P(X=r)=\binom{n}{r}\cdot p^r\cdot(1-p)^{n-r}$$

Example: Ms. Carey rolls her six-sided Bicycle die ten times. Find the probability that she rolls "bicycle" exactly three times.

Binomial with $n = 10$ and $p=\frac{1}{6}$.

$$P(X=3)=\binom{10}{3}\cdot\left(\tfrac{1}{6}\right)^3\cdot\left(\tfrac{5}{6}\right)^7\approx 0.155$$

- Probabilities for binomial distributions can also be found using the GDC. For $X \sim B(n,\ p)$ there are two buttons to choose from:
  - to find $P(X=r)$ use binompdf(*n*, *p*, *r*)
  - to find $P(X\le r)$ use binomcdf(*n*, *p*, *r*)

Example: Ms. Carey rolls her six-sided Bicycle die ten times. Find the probability that she rolls "bicycle" no more than three times.

Binomial with $n = 10$ and $p=\frac{1}{6}$.

$$P(X\le 3)=\text{binomcdf}\left(10,\ \tfrac{1}{6},\ 3\right)\approx 0.930$$

- Note that the binomcdf button on the GDC only gives probabilities for X less than or equal to r. If you need to find $P(X > r)$ or $P(X \geq r)$ you must use complementary events.

  Example: Ms. Carey rolls her six-sided Bicycle die ten times. Find the probability that she rolls "bicycle" at least three times.

  Binomial with $n = 10$ and $p = \frac{1}{6}$.

  $P(X \geq 3) = 1 - P(X \leq 2) = 1 - \text{binomcdf}\left(10,\ \frac{1}{6},\ 2\right) \approx 0.224$

  Example: Five percent of the eggs produced on a farm are brown. Find the probability that, in a case of 240 eggs, there are more than ten brown eggs.

  Binomial with $n = 240$ and $p = 0.05$.

  $P(X > 10) = 1 - P(X \leq 10) = 1 - \text{binomcdf}\left(240,\ 0.05,\ 10\right) \approx 0.658$