MATH 460: Probability & Statistics — January_2019

Ch 1: Basic Probability ∪ Ch 5 Sampling Theory

Class 2 Notes — **Dr. Basilio**

Southern California Institute of Technology

Wed Jan_9 ∪ Thurs Jan_10

* * *

Guiding Question(s)

(1) What are different ways of counting combinations of things?

(2) What are different ways to organize or visualize data?

(3) What does it mean to be at the "center" of a data set?

Fundamental Principle of Counting

We want to be able to count the total number of ways we can combine things. Here's a few examples to illustrate the main points:

Example 1: Fundamental Principle of Counting

If we want to flip a coin and then roll a die, how many different possibilities are there? With a little work we could simply list out all the options: H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, and T6. (Where H stands for heads and T stands for tails.) There are 12 outcomes, which makes sense when you think of the two sided coin and the six sided die: $2 \times 6 = 12$.

Example 2: Fundamental Principle of Counting

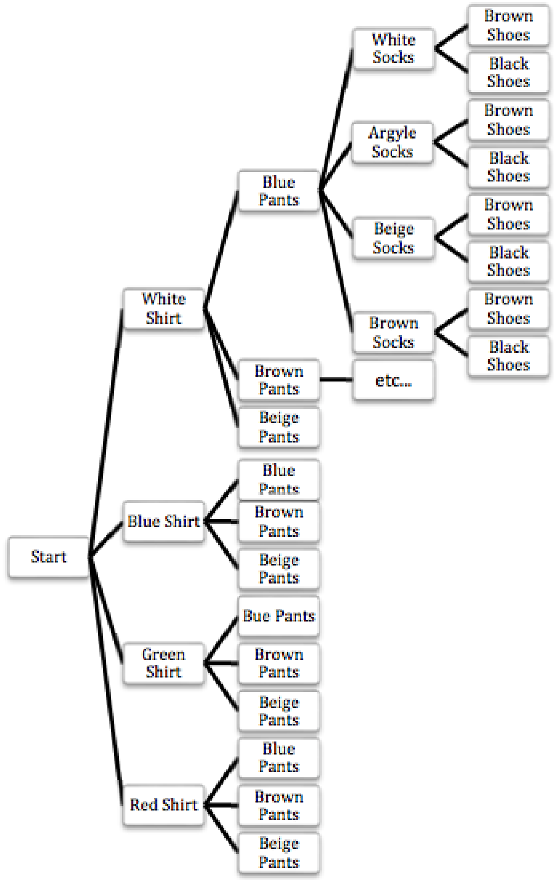
Imagine you want to get dressed in the morning and you lay out on your bed 4 shirts, 3 pairs of pants, 4 pairs of socks, and in addition you have 2 pairs of shoes available. If we assume that all of the items you laid out match well, how many different ways can you put a complete outfit together?
The number of ways you can put together a complete outfit is: $4 \times 3 \times 4 \times 2 = 96$. Thus, there are 96 different outfits possible. Enough choices to make you late for work!!

Theorem 1: Fundamental Principle of Counting

When we have independence and use the Fundamental Counting Principle, we can think of each level of choice as a column in a slot machine. Each category fits in its own column in the slot machine and the choices can revolve independently. Or we think of it as filling in the blanks at each step with the number of choices:

$$\underset{\text{\# of choices of item 1}}{\underline{\hspace{3cm}}} \times \underset{\text{\# of choices of item 2}}{\underline{\hspace{3cm}}} \times \underset{\text{\# of choices item 3}}{\underline{\hspace{3cm}}} \times \cdots = \underset{\text{TOTAL \# of ways to combine}}{\underline{\hspace{3cm}}} \tag{1}$$

Key Point: the FPoC works *with* or *without* replacing the items.



Activity 1: FPoC

(a) How many different 4-digit PINs are there? Recall that there are ten digits: 0–9.

(b) You have bought 4 new books and you want to display them on your bedroom shelf. You have A: *Alice in Wonderland*, B: *Bleak House*, C: *Crime and Punishment*, D: *Don Quixote of La Mancha*. How many different ways can we actually organize these 4 books?

(c) In Georgia car license plates have three digits followed by three letters. Unless stated assume you can repeat letters/numbers. How many are possible:

  (i) With no restrictions?

  (ii) With no repeated digits or letters?

  (iii) That start with zero?

  (iv) That do NOT start with zero?

  (v) That have the word "DOG"

Definition 1: Factorial

The **factorial symbol** of a positive whole number n is defined to be the product of n with all the numbers decreasing by one down to 1. It is denoted by $n!$ for brevity. This means:

$$\boxed{n! = n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 3 \times 2 \times 1}$$

We also define $0!$ to be 1, that is $0! = 1$.
By the Fundamental Principle of Counting, $n!$ is the number of ways of arranging n objects in a line *without replacement*.

Permutations

Suppose that we are given n distinct objects and wish to arrange r of these objects in a line. This is *without repeating the objects!* Notice also that because we are arranging the objects in a line that ORDER MATTERS!

Since there are n ways of choosing the 1st object, and after this is done, $n-1$ ways of choosing the 2nd object, ..., and finally $n-(r-1)$ ways of choosing the rth object, it follows by the fundamental principle of counting that the number of different arrangements, or **permutations** as they are often called, is given by

$$ {}_nP_r = n \cdot (n-1) \cdot (n-2) \cdots (n-(r-1)) \tag{2}$$

The notation ${}_nP_r$ is the number of different permutations of n objects taken r at a time.
When $r = n$, that is, when we choose n objects and arrange them, there are

$$ {}_nP_n = n \cdot (n-1) \cdot (n-2) \cdots (n-(n-1)) = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n! \tag{3}$$

Theorem 2: Permutations

When selecting n objects and order them r at a time *without replacement*, that is, permutation of n objects of length r, we have

$$ {}_nP_r = \frac{n!}{(n-r)!} = n \cdot (n-1) \cdot (n-2) \cdots (n-(r-1)) \tag{4}$$

Example 3: Permutations

The number of different arrangements, or permutations, consisting of 3 letters each that can be formed from the 7 letters A, B, C, D, E, F, G is

$$ {}_7P_3 = 7 \cdot 6 \cdot 5 = 210$$

or we can calculate it using the formula from the theorem:

$$ {}_7P_3 = \frac{7!}{4!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = 210$$

Activity 2: Permutations

Calculate the following:

(a) $10!$

(b) ${}_8P_5$

(c) ${}_4P_4$

Activity 3: Permutations

(a) In how many ways can 10 people be seated on a bench if only 4 seats are available?

(b) Castel and Joe are planning trips to three countries this year. There are 7 countries they would like to visit. One trip will be one week long, another two days, and the other two weeks. How many possibilities are there?

Combinations

In a permutation we are interested in the order of arrangement of the objects. For example, abc is a different permutation from bca. In many problems, however, we are interested only in selecting or choosing objects *without regard to order AND without replacement*. Such selections are called **combinations**. For example, abc and bca are the same combination.
The total number of combinations of r objects selected from n (also called the combinations of n things taken r at a time) is denoted by ${}_nC_r$ or $\binom{n}{r}$.

Theorem 3: Combinations

When computing the n combinations of r objects, we have

$$ {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{{}_nP_r}{r!} \tag{5} $$

Example 4: Combinations

The number of ways in which 3 cards can be chosen or selected from a total of 8 different cards is

$$ {}_8C_3 = \binom{8}{3} = \frac{8!}{3!5!} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56. $$

Activity 4: Combinations

Calculate the following:

(a) ${}_{12}C_{10}$

(b) ${}_7C_7$

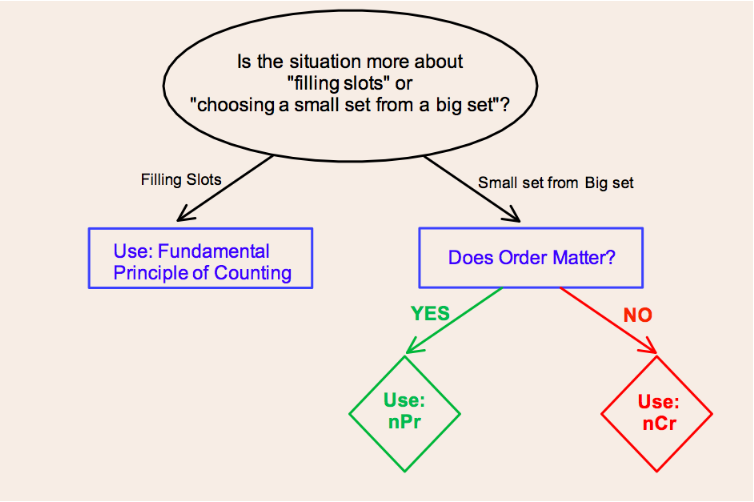
Activity 5: Permutations and Combinations

(a) In how many ways can 10 objects be split into two groups containing 4 and 6 objects, respectively?

(b) In how many ways can a team of 17 softball players choose three players to refill the water cooler?

(c) I have 50 math books in my collection and want to donate 5 to the local library. In how many ways can I do this?

We learned 3 different techniques to help us count the number of ways: Fundamental Principle of Counting, Permutations, and Combinations. This flow chart helps you know how to choose the right technique:



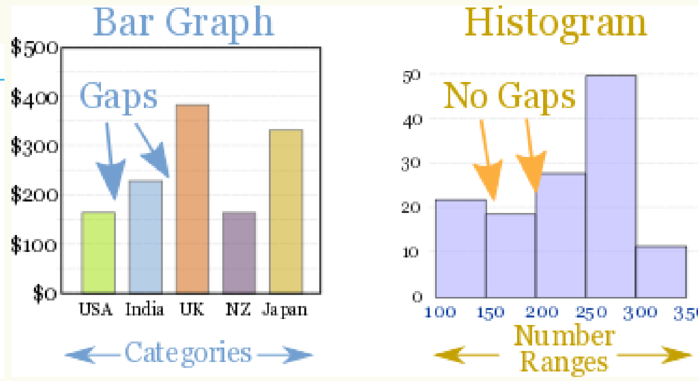
Organizing and Visualizing Data

Definition 2: Data

- **FREQUENCY DISTRIBUTION:** It is useful to organize or group the raw data. We arrange the data into classes or categories and determine the number of individuals belonging to each class, called the **class frequency**. When arranged as a table it is called a **frequency distribution** or **frequency table**.
- **RELATIVE FREQUENCY DISTRIBUTION:** If we recorded the relative frequency or **percentage** rather than the number in each class/category, the result would be a **relative, or percentage, frequency distribution**.
- **BAR GRAPHS and HISTOGRAMS:**

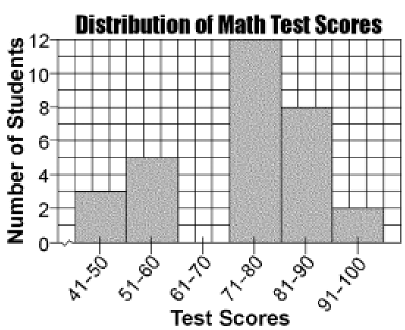
Type of smartphone	Tally	Frequency	Relative Frequency
Android	卌 卌 II	12	0.3
Window Phone	卌 III	8	0.2
Iphone	卌 卌 卌	15	0.375
Amazon's fire phone	卌	5	0.125
		Sum = 40	

Frequency distribution of types of smartphones owned



Activity 6: Frequency Distributions

The graph below shows the distribution of scores of 30 students on a mathematics test.



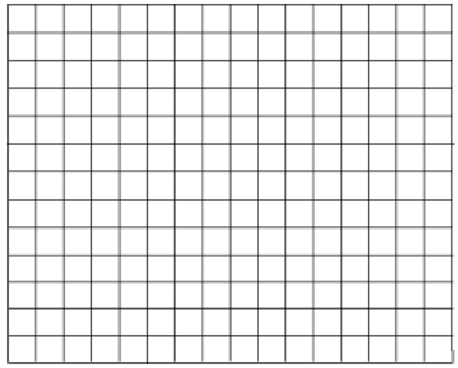
Complete the frequency table below using the data in the frequency histogram shown.

Test Scores	Frequency
91–100	
81-90	
71-80	
61-70	
51-60	
41-50	

Activity 7: Frequency Distributions

The scores on a mathematics test were 70, 55, 61, 80, 85, 72, 65, 40, 74, 68, and 84. Complete the accompanying table, and use the table to construct a frequency histogram for these scores.

Score	Tally	Frequency
40-49		
50-59		
60-69		
70-79		
80-89		

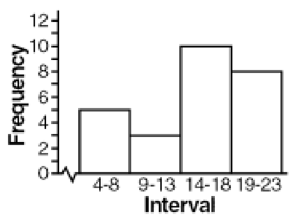


(a)

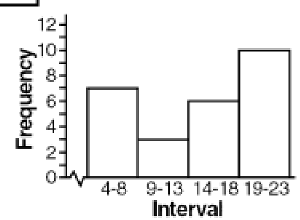
Which one of the following histograms represents the data in the table below?

Interval	Frequency
4-8	8
9-13	3
14-18	10
19-23	5

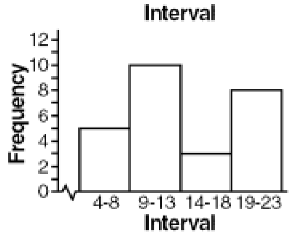
A)



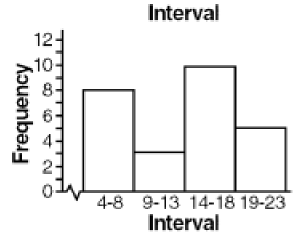
C)



B)



D)



(b)

Measurements of Central Tendency

There are different approaches for measuring the "center", so we have different definitions for those different approaches.

Definition 3: Mean-Median-Mode-Range

- **MEAN:** the mean is computed by adding all of the numbers in the data together and dividing by the number elements contained in the data set. This is also called the **average**.
- **MEDIAN:** The median of a data set is dependent on whether the number of elements in the data set is odd or even. First re-order the data set from the smallest to the largest then if the number of elements are odd, then the Median is the element in the middle of the data set. If the number of elements are even, then the median is the average of the two middle terms.
- **MODE:** The mode for a data set is the element that occurs the most often. It is not uncommon for a data set to have more than one mode. This happens when two or more elements occur with equal frequency in the data set. A data set with two modes is called **bimodal**. A data set with three modes is called **trimodal**.

Less common, but also useful are:

- **MIDRANGE:** the mean of the maximum value and the minimum value.

Activity 8: Mean-Median-Mode-Range

(a) Consider the data set $S = \{2, 5, 9, 3, 5, 4, 7\}$. Compute the mean.

(b) Consider the data sets $A = \{2, 5, 9, 3, 5, 4, 7\}$ and $B = \{2, 5, 9, 3, 5, 4\}$. Compute the median of each data set. (Don't forget to re-order the data first!)

(c) Consider the data sets $A = \{2, 5, 9, 3, 5, 4, 7\}$, $B = \{2, 5, 2, 3, 5, 4, 7\}$, $C = \{2, 5, 2, 7, 5, 4, 7\}$. Compute the mode(s) of each data set.

(d) Consider the data set $S = \{2, 5, 9, 3, 5, 4, 7\}$. Compute the range and the midrange of the data set.

Which of these measures of center are sensitive to **outliers** (that is, one extreme value)?

Theorem 4: Properties of Measures of Center

- **MEAN:**
  - Only for Quantitative data
  - Very sensitive to outliers
  - Uses all data values to compute
- **MEDIAN:**
  - Only for Quantitative data
  - NOT sensitive or resistant to outliers
  - Doesn't use all data values to compute
- **MODE:**
  - Used for Quantitative and Qualitative data
  - Can have no mode or more than one (bimodal, trimodal, etc)

Calculator Tips

Entering data

First, you need to enter your data points in the list editor. To do this,
1. Select $\boxed{\text{STAT}}$ > $\boxed{\text{EDIT}}$.
You should see headings L1, L2, and L3.
2. List the input values from your data points under L1.
If you want to keep this data and add another data set, you can use L2, etc.

Basic Statistics

1. Press $\boxed{\text{STAT}}$ then select $\boxed{\text{CALC}}$ from the top menu.
2. Select $\boxed{\textbf{1-Var Stats}}$ and press $\boxed{\text{ENTER}}$.
3. Enter the name of list that includes the desired data (e.g., L1).
4. Select $\boxed{\textbf{Calculate}}$ and press $\boxed{\text{ENTER}}$ to view descriptive statistics.
TIP: Press down arrow to view additional statistics that don't fit on the initial screen.

Measurement of Dispersion and Variation

We want to know how to compare data that is very "spread out" to data that is "clumped together," or simply put how data is varied.

One basic way to measure how data is varied is with the:

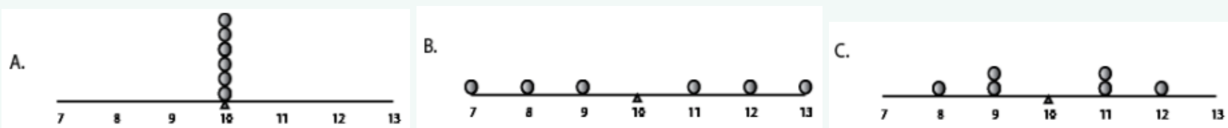
- **RANGE:** the difference between the maximum value and the minimum value.

Example 5: Dot Plots

Each of these distributions (or dot plots) has a mean and median of 10, but each distribution has a different amount of spread. Based on looks, order the distributions from least amount of variability to most amount of variability.

Using Range: Least Spread .... Most Spread

Using Intuition: Least Spread .... Most Spread



Definition 4: Standard-Deviation

To get a better measure of spread, we want to use all the pieces of data. To do this, we need to compare each data to the "center"–so we compare each data to the mean. This is called **deviation**, notation is $x - \bar{x}$. Then we basically take the average of the deviations. But the deviations can be positive or negative, and we don't really care about which side of the mean we are on, *only how far away from the mean!* So to make each deviation positive we square them.

The notation for this is given by:

$$\boxed{\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}} \tag{6}$$

- x: a value from a data set
- $\bar{x}$: mean
- n: number of values in a data set
- Σ: "to sum or add" (Capital Greek letter "Sigma")
- σ: standard deviation (lower-case Greek letter "sigma")

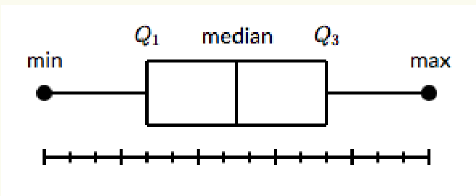
Data Set 1			
x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
12.6	15	-2.4	5.76
15.1	15	0.1	0.01
11.2	15	-3.8	14.44
17.9	15	2.9	8.41
18.2	15	3.2	10.24
$\frac{\sum(x-\bar{x})^2}{n}$			7.772
Standard Deviation: $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$			≈ 2.79

The data set with the larger standard of deviation has a larger more spread out range of values.

If many of the data values are close to the mean, then the data would have a relatively small standard deviation. This would tell you that the data is not very spread out.

Definition 5: Five-Number-Summary

- **RANGE:** The range for a data set is the difference between the largest value and smallest value contained in the data set. First reorder the data set from smallest to largest then subtract the first element from the last element.
- **FIVE NUMBER SUMMARY:** is the minimum, first quartile, median, third quartile, and maximum.
- **Box-Whisker Plot:** displays the five-number summary of a set of data. In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.



Activity 9: Five-Number-Summary

(a) Find the five number summary, and draw a Box-Whisker plot for $S = \{15, 25, 20, 29, 29, 36, 29, 15, 26, 28, 24, 25\}$.

(b) Find the standard deviation for the set from problem 4.