Notes for Lesson 1

Two Way Frequency Table

We can fill out frequency tables based on surveys that our conducted. For instance, let's say you conducted a survey in the class to see how many students have Android phones, iPhones, or other (which includes no phone). If you surveyed the 11th and 12th grades you might get the following information.

	Has an Android Phone	**Has an iPhone**	**Other**	**Total**
11th Grade	34	27	13	74
12th Grade	29	31	12	72
Total	63	58	25	146

Sometimes we will have to fill out data that in a two way frequency table that is not fully complete. We can place data in empty cells that makes the totals of the columns and rows add up.

For instance, consider the following example:

Example 1 - Two Way Frequency Tables

1. Elena surveyed her schoolmates about whether they have an unlimited data phone plan and whether they believe they spend a lot of time on their phones. Complete the missing entries in this two-way table.

	A Lot of Time on the Phone	**Not a Lot of Time on the Phone**	**Total**
Unlimited Data		10	
Don't Have Unlimited Data	32		
Total	84	53	

We can start filling this out by focusing on the bottom row and calculate the total:

	A Lot of Time on the Phone	**Not a Lot of Time on the Phone**	**Total**
Unlimited Data		10	
Don't Have Unlimited Data	32		
Total	84	53	137

We then can figure out the first cell by subtracting 32 from 84 to get 52.

	A Lot of Time on the Phone	Not a Lot of Time on the Phone	Total
Unlimited Data	52	10	
Don’t Have Unlimited Data	32		
Total	84	53	137

We can then subtract 10 from 53 to get 43 and place that in the appropriate cell.

	A Lot of Time on the Phone	Not a Lot of Time on the Phone	Total
Unlimited Data	52	10	
Don’t Have Unlimited Data	32	43	
Total	84	53	137

Finally, we can add our rows across to get the respective row totals.

	A Lot of Time on the Phone	Not a Lot of Time on the Phone	Total
Unlimited Data	52	10	62
Don’t Have Unlimited Data	32	43	75
Total	84	53	137

There we go. In another example, we may have to figure out if there is an association between the variables represented in the table.

Example 2 - Two Way Frequency Tables

Diego is investigating whether there is an association between wearing sneakers and participating on an athletic team. This table summarizes the data he collected.

	On Athletic Team	Not on Athletic Team	Total
Wear Sneakers	12	18	30
Don’t Wear Sneakers	24	36	60
Total	36	54	90

Is there evidence of an association between the variables? Explain How you know.

Answer: No, 40% of the students who wear sneakers are on the athletic team (12/30 = 0.4) which is the same percentage of students who don’t wear sneakers and are on the athletic team.

Two Way Relative Frequency Tables

Consider the following *two way frequency table* representing people who have kids and pets, don't have kids but have pets, have kids who don't have pets, and doesn't have kids or pets:

	Has Kids	**Doesn't Have Kids**	**Total**
Has Pets	45	21	66
Doesn't Have Pets	34	23	67
Total	79	44	133

We can create a *two way relative frequency table* by figuring out the percentage of the total each group represents.

	Has Kids	**Doesn't Have Kids**
Has Pets		
Doesn't Have Pets		

We can fill out this table with percents of the total:

	Has Kids	**Doesn't Have Kids**
Has Pets	33.8%	15.8%
Doesn't Have Pets	25.6%	17.3%

Correlation

Consider the following bivariate data from a sample of 15 adults who were asked about their age and annual salary.

Age	**18**	**22**	**25**	**30**	**33**	**38**	**41**	**43**	**45**	**47**	**53**	**58**	**61**	**64**	**64**
Salary (in 1000s)	31	62	45	71	67	83	74	77	81	97	99	101	45	120	67

Create a scatterplot for the data using Desmos.

In a separate cell type in $y_1 \sim mx_1 + b$.

This will give you the line of best fit, which is:

$y = 0.918046x + 35.3743$

Which a correlation coefficient, *r*, of 0.5873. This would be classified as a moderate relationship (between age and salary).

We have some more bivariate data from a sample of 10 adults who were asked about their number of years of education and salary.

# of years of education	13	19	4	16	14	17	20	13	10	13
Salary (in 1000s)	25	88	10	44	38	55	74	31	20	50

Create a scatterplot with Desmos.

Then in a separate cell type in $y_1 \sim mx_1 + b$.

You'll see that the line of best fit is: y = 4.63712x - 20.9559.

The correlation coefficient is r = 0.8856.

Is there a relationship between the number of years of education and the salary.

Yes.

Is the relationship positive or negative?

Positive.

Is this relationship weak or strong?

Strong.

In these two examples we are studying the linear relationship between the two variables.

We can refer to relationship strength as strong, moderate, or weak. We can refer to the direction of the relationship as positive, negative, or no association.

Expectations of the Type of Relationship

- Number of hours studied and the grade on the final exam. We would expect this relationship to be positive.
- Age of a maple tree and the diameter of the trunk. We would expect this relationship to be positive.
- Size of an automobile and the gas mileage it gets. We would expect this relationship to be negative.
- The IQ of a person and the year they were born. We would expect to have no association (none).
- The average temperature and the distance from the equator. We would expect this to have a negative relationship.
- Grade point average and shoe size. We would expect this to have no association.

Pearson Correlation coefficient

We will be using *r* to represent the correlation coefficient for our samples. (Fun fact: For a population the Greek letter ρ is used. A measure from an entire population is called a parameter while the measure from a sample population is called a statistic.) We typically study representative samples of a population (statistics). There is an equation to find *r* but we will be using Desmos (technology). You will also be able to use Desmos on all math questions on the digital SAT so it is in your best interest to learn it.

r will always be no less than -1 and no greater than 1.

$-1 \leq r \leq 1$

If r is negative we have a negative association between our variables. If r is positive we have a positive relationship between our variables. When r has a value of 0 have no *linear correlation*.

- $-0.5 \leq r \leq 0.5$ (weak correlation)
- $-0.8 \leq r \leq -0.5$ (moderate correlation)
- $0.8 \leq r \leq 0.5$ (moderate correlation)
- $-1 \leq r \leq -0.8$ (strong correlation)
- $1 \leq r \leq 0.8$ (strong correlation)

Note: These values depend on the source. For instance, the link from The University of West Georgia has slightly different values. I'm not going to get hung up on which values you use so long as you support your reasoning.