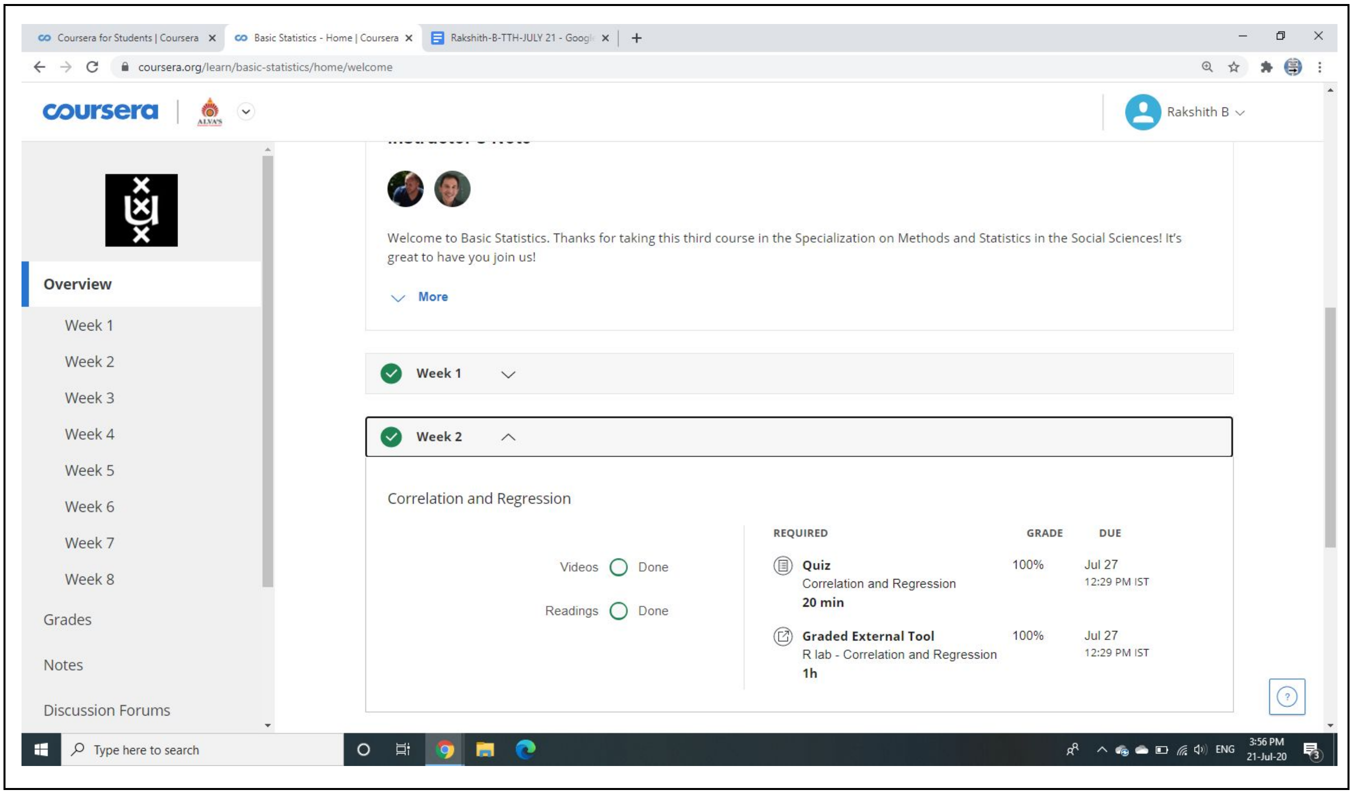
REPORT JULY 21

Date:	21 JULY 2020	Name:	Rakshith B
Course:	Coursera	USN:	4AL16EC409
Topic:	Basic statistics	Semester & Section:	6th SEM B
Github Repository:	Rakshith-B		

Image of the Session



The first two videos in this module discuss the concept of **correlation**. In the first video, we'll talk about how we can display the correlation between two variables using tables and graphs. First we'll look at categorical variables and discuss **contingency tables**. In a next step we look at how we can best display the relationship between two quantitative variables. Here we'll introduce the **scatterplot**.

In the second video we'll discuss Pearson's **r** - one of the most frequently used measures of correlation. It is an appropriate measure if the variables under analysis are measured on a quantitative level and if they are linearly related to each other. The Pearson's r expresses the direction and strength of the correlation. You'll learn how to interpret the Pearson's r and how to compute it yourself.

Many people like eating chocolate. But most people are somewhat cautious with their chocolate consumption. Because it might well be the case that eating a lot of chocolate increases your body weight.

In this video today, I'll talk about how we can display a relationship between two variables using tables and using graphs. This can be very useful to help you discover if two variables are correlated or not.

Let us investigate the relationship between eating chocolate and body weight a little bit further.

Suppose I have selected 200 female students at my university, who are all one meter seventy tall. This way, height is a constant and cannot account for differences in body weight or chocolate consumption

I asked the students to report their body weight and their weekly chocolate consumption. They could choose between the categories less than 50 kilograms, 50 to 69 kilograms, 70 to 89 kilograms, and 90 kilograms or more.

They could indicate their chocolate consumption by choosing less than 50 grams per week, between 50 and 150 grams per week, and more than 150 grams per week.

Here are the results. What you see here is a contingency table. A contingency table, enables you to display the relationship between two ordinal or nominal variables. It is similar to a frequency table. But the major difference is that a frequency table always concerns only one variable, whereas a contingency table concerns two variables.

In our study we have two variables. Body weight and chocolate consumption.

The table shows that we have 33 individuals with a body weight of less than 50 kilograms. 27 of them eat less than 50 grams of chocolate per week.

You can also see that 90 individuals eat between 50 and 150 grams of chocolate per week. 7 of them weigh 90 kilograms or more.

In this form, the table does not tell you much yet about a correlation between the two variables, because the columns and rows contain different numbers of cases. It provides more insight, when you compute percentages. And in this case, we compute column percentages. This means that for every cell, we compute a percentage of cases in that cell, compared to the total number of cases in the corresponding column. So for instance, in this cell, we have 24 cases. The total number of cases in the column is 60. So the percentage is 24 divided by 60 times 100. That equals 40%. When we do that for every cell, this is the result.

We can also express these percentages as proportions. 45% then becomes 0.45. 38% becomes 0.38. We call these proportions, conditional proportions because the formation is conditional on another variable. In this case, that variable is chocolate consumption.

We can also ignore the information we have about chocolate consumption, and use the count in the margin of the table. These are marginal proportions. For instance, 33 divided by 200. That makes 0.17. And this proportion tells you that a proportion of 0.17, that's the same as 17%, of the respondents in the study weighs less than 50 kilograms.

Also, of those who eat less than 50 grams of chocolate, 45% weigh less than 50 kilograms, while those who eat more than 150 grams of chocolate. Only 2% weigh less than 50 kilograms.

These percentages show that those who eat more chocolate are also more likely to weigh more, and those who eat less chocolate are also more likely to weigh less. In other words, the percentages show that there is a correlation between chocolate consumption and body weight.

A contingency table is useful for nominal and ordinal variables. But not for quantitative variables. For quantitative variables a scatterplot is more appropriate. Suppose that instead of providing categories, I asked the 200 women to give me their exact body weight. For instance, 65 or 72 kilograms. Suppose I also ask them to tell me how much chocolate they eat every week. That could be, for instance, 64 or 99 grams per week.

Now, I have much more precise information than before. And the best way to display the relationship between the quantitative variables, chocolate consumption and weight, is with a scatterplot.

To make a scatterplot we draw two lines which we call axes. We call the horizontal axis the X axis. Here we display the independent variable. And the vertical axis is called the Y axis, which we use to represent the dependent variable.

If there is no distinction between dependent and independent variable, the placement on the Y axis and X axis is a matter of choice.

In our case, the independent variable is chocolate consumption, and the dependent variable is body weight.

Imagine that our study shows that the lowest amount of chocolate eaten is equal to zero grams per week. And the highest amount is 700 grams per week. We displayed these values on the x axis. Similarly, the minimum value when it comes to body weight is 40 kilograms, and a maximum value is 110 kilograms.

Next, we display every individual in this figure. For instance, here is one person that eats 49 grams of chocolate per week and weighs 65 kilograms. Another individual eats 134 grams of chocolate and weighs 73 kilograms. That's here.

We do that for all individuals in our sample. And voila, there's the scatterplot. The scatterplot shows you at a glance that there is a relationship between chocolate consumption and body weight. The more chocolate you eat, the higher your body weight.

What have you learned? Not that chocolate consumption and body weight are correlated. I think most of you were already aware of that relationship. What you have learned in this video is that we can display relationships between two variables by means of tables and by means of graphs. When the variables in the study are measured on the nominal, or ordinal level, we use a contingency table. And, when they are measured on a quantitative level, we use a scatterplot.

This scatterplot shows the relationship between chocolate consumption and body weight.

On the horizontal X axis we see the independent variable chocolate consumption, measured in consumed grams of chocolate per week.

On the vertical Y axis, we have the dependent variable body weight, measured in kilograms. The case study here are 200 female students, who are exactly one meter 70 tall. They are represented by the 200 dots in this graph.

The scatterplot shows at a glance that there is a strong correlation between the two variables. The more chocolate someone eats, the larger the body weight.

But how strong is this correlation? We will now turn to one of the most often used measures of correlation, the Pearson's r. One of the most important advantages of the Pearson's r, is that it expresses the direction and strength of the linear correlation between two variables with one single number.

Conclusion, we have a strong, positive and linear relationship here. However, variables could also be correlated in different ways.

In this graph, we see a rather strong positive and linear relationship between the variables x and y, just like in the example with chocolate consumption and body weight

But in this graph, we have a rather strong negative linear correlation. The line goes down that indicates that when variable x goes up, variable y goes down.

In this graph, we also see a positive linear relationship. However, it is much less strong than the previous ones. We know that because the individual cases are much further removed from the line. This is a perfect negative linear correlation. It is perfect because all cases lie exactly on the line.

But the correlation between two variables need not be linear. In this graph, we also see a relationship between the variables x and y. However, the line that best represents the relationship between the two variables is not straight. Instead, it is a U-shaped line. We call this a curvilinear relationship.

A scatter plot helps us to broadly assess whether a correlation is strong or weak. But it does not tell us exactly how strong the relationship is.

Pearson's r is a measure that can show us exactly that. More specifically, the Pearson's r tells us the direction and exact strength of the linear relationship between two qualitative variables. A positive Pearson's r indicates that a correlation is positive, and a negative correlation indicates that it is negative.

The size of R expresses how tightly the observations are clustered around the imaginary best fitting straight line through the cloud of the data points. Pearson's r is always a number between -1 and 1. Minus one refers to a perfect negative correlation, and plus one to a perfect positive correlation. And zero means that there is no correlation at all.

So, how do we compute a Pearson's r? Imagine that the study on chocolate consumption and body weight was not based on 200, but on only four individuals.

You can see that every combination of values on the two variables becomes a circle in the graph. This woman consumes 200 grams of chocolate per week and weighs 70 kilograms. She's represented by this circle. The other three circles represent the other individuals you see in the data matrix. To compute a Pearson's r we need this formula.

What does it mean? Well, first we change all original scores to z-scores. In other words, we standardize the values. The reason is that we want the Pearson's r to be a number between minus one and one. If we don't standardize, the measure of correlation will be expressed according to the original metrics.

So, first, we compute the mean for both variables. This results in the value 162.5 for variable x, that's chocolate consumption, and 71.25 for variable y, that's body weight.

Then we compute the standard deviations for both variables. The results are 110.9 for x and 18.4 for y.

We the get the z-scores by applying this formula to every case. We subtract the mean from every value, and then divide it by the standard deviation. So, 50-162.5/110.9=-1.01.

We do that for every value of the independent variable, chocolate consumption. And for every value of the dependent variable, body weight. In the next step, we compute the products of every z-score on x for every z-score on y. So for the first case, that's -1.01, times -1.15 equals 1.17.

These are the results for the other three cases.

To get this part of the formula, we add up all these scores, that's 2.78. To finish, we have to divide it by n minus 1. N = 4, so in our case, n - 1 means 4 - 1, that's 3. The Pearson's r is 2.78 divided by 3, that equals 0.93. What does this mean? It means that there is a strong positive linear relationship between chocolate consumption and body weight.

One important note, though, you can always compute a Pearson's r, even if the relationship is not linear. Therefore, it is very important that before you compute a Pearson's r, you can always check the scatterplot first to see if your variables are linearly related.

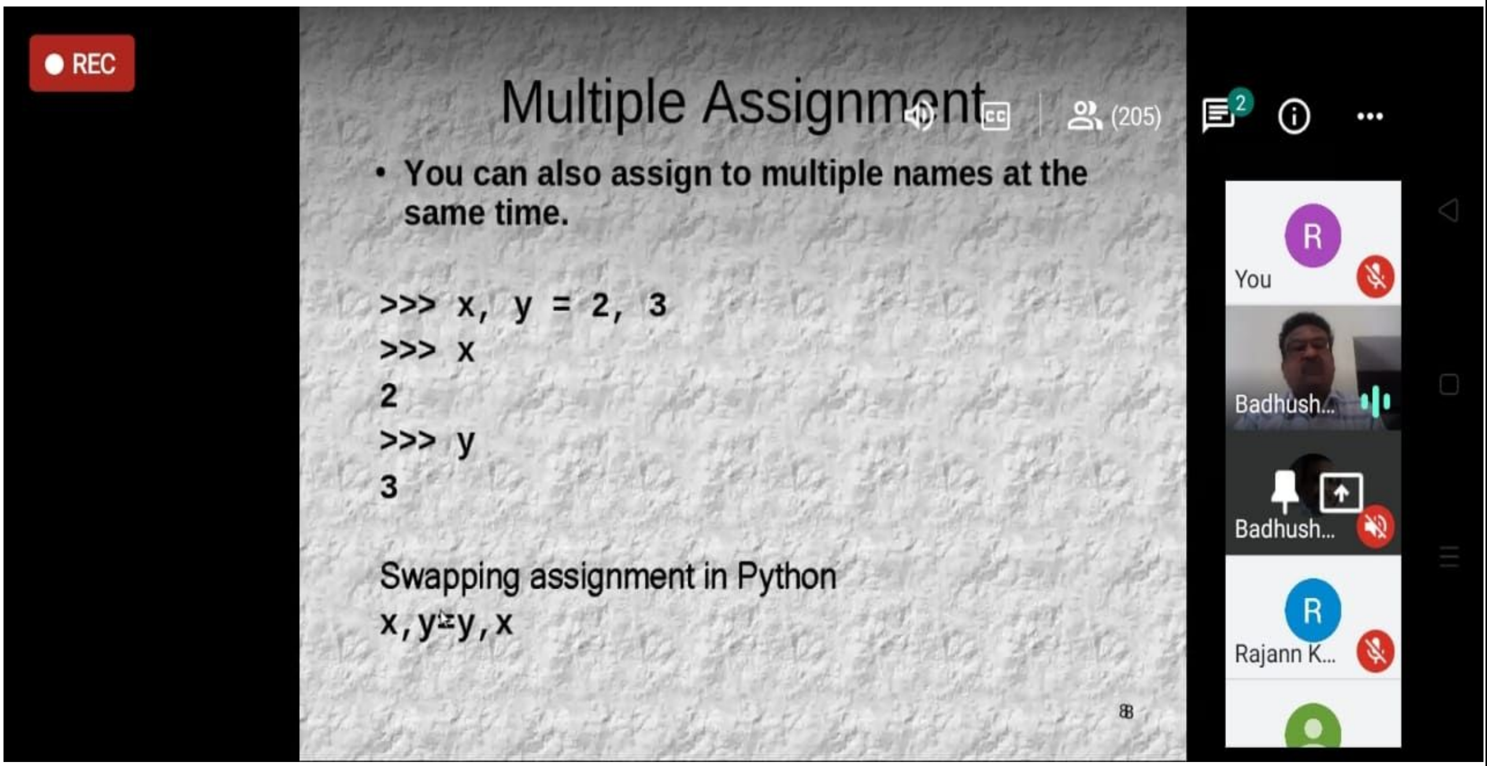
If not, do not compute a Pearson's r, because it doesn't tell you much about the relationship between your variables. This scatterplot, for instance, shows that there's a strong curvilinear relationship between x and y. If you compute the Pearson's r you will get a very low value, -0.15. This doesn't tell you that there is a weak correlation, it only tells you that there is a weak linear correlation.

It is rather easy to compute a Pearson's r with only four cases. However, you can probably imagine that it becomes an almost impossible task when you have, say, 200 cases. Luckily, every statistical computer program can compute a Pearson r in no time. Nevertheless, it is important for you to understand what the Pearson r exactly tells you. And it's also important to understand what the formula means.

It helps you to better understand how variables are related. And it might help you to decide for yourself how much chocolate to eat every week.

Date:	21 JULY 2020	Name:	Rakshith B
Course:	Pythonic coding	USN:	4AL16EC409
Topic:	Python	Semester & Section:	6th SEM B
Github Repository:	Rakshith_B_colab		

Image of the Session :




```
#Python Program to Add Two Numbers getting through key board
# sum of two nos
num1 = int(input("Enter first no"))
num2 = int(input("Enter second no"))

# Adding the two numbers
sum = num1 + num2
```

Display the sum

```
print('The sum of {0} and {1} is {2}'.format(num1, num2, sum))
```

Python program to check if the input year is a leap year or not

```
# To get year (integer input) from the user
year = int(input("Enter a year: "))

if ((year % 4) == 0 and (year % 100)  != 0) or ((year % 400) == 0):
            print("{0} is a leap year".format(year))
else:
            print("{0} is not a leap year".format(year))
```

Python Program to Generate a Random Number

```
# Program to generate a random number between 0 and 9
# import the random module
import random
print(random.randint(0,9))
```

Python Program to Convert Kilometers to Miles

```
# To take kilometers from the user, uncomment the code below
kilometers = int(input("Enter value in kilometers"))

# conversion factor
conv_fac = 0.621371
```

```
# calculate miles
miles = kilometers * conv_fac
print('%0.3f kilometers is equal to %0.3f miles' %(kilometers,miles))
```

Python Program to Solve Quadratic Equation

```
# Solve the quadratic equation ax**2 + bx + c = 0
# importing  complex math module
import cmath

 # To take coefficient input from the users
a = float(input('Enter a: '))
b = float(input('Enter b: '))
c = float(input('Enter c: '))

# calculate the discriminant
d = (b**2) - (4*a*c)

# find two solutions
sol1 = (-b-cmath.sqrt(d))/(2*a)
sol2 = (-b+cmath.sqrt(d))/(2*a)

print('The solution are {0} and {1}'.format(sol1,sol2))
```

Python Program to find prime or not using function

```
def test_prime(n):
```

```
    if (n==1):
        return False
    elif (n==2):
        return True;
    else:
        for x in range(2,n):
            if(n % x==0):
                return False
        return True
no=int(input("Enter the number"))
if (test_prime(no)) is True :
    print(" {0} is a prime no".format(no))
else:
     print(" {0} is not a prime no".format(no))



#CODING
"""Calculator program"""

loop = 1 # 1 means loop; anything else means don't loop.
choice = 0 # This variable holds the user's choice in the menu

def add(a,b):
    return a+b
def sub(a,b):
    return a-b
def mul(a,b):
    return a*b
def div(a,b):
```

```
    return a/b

while loop == 1:
    # Print what options you have
    print ("Welcome to calculator.py")
    print ("your options are:")
    print (" ")
    print("1) Addition")
    print("2) Subtraction")
    print("3) Multiplication")
    print("4) Division")
    print("5) Quit calculator.py")
    print(" ")
      try:
        choice = int(input("Choose your option: "))
    except:
        print('please enter a valid number for option')
    print(" ")
    print(" ")
    if choice == 1:
        x = int(input("Enter 1st no: "))
        y = int(input("Enter 2nd no: "))
        print("The answer is ",add(x,y))

   elif choice == 2:
        x = int(input("Enter 1st no: "))
        y = int(input("Enter 2nd no: "))
        print("answer is ",sub(x,y))
```

```
  elif choice == 3:
        x = int(input("Enter 1st no: "))
        y = int(input("Enter 2nd no: "))
        print("answer is ",mul(x,y))

    elif choice == 4:
        x = int(input("Enter 1st no: "))
        y = int(input("Enter 2nd no: "))
        print("answer is ",div(x,y))

    elif choice == 5:
        loop = 0

   else:
        print("please choice a valid option from 1 to 5")
        choice=0
print ("Thank-you for using calculator.py!")
```