



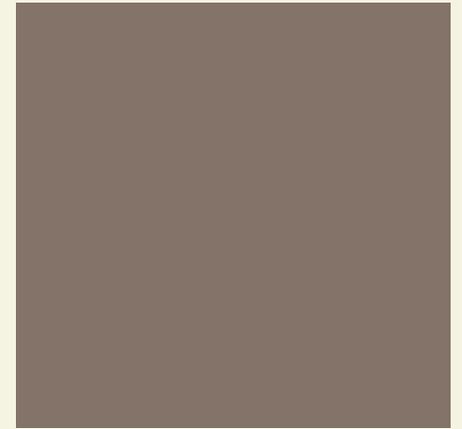
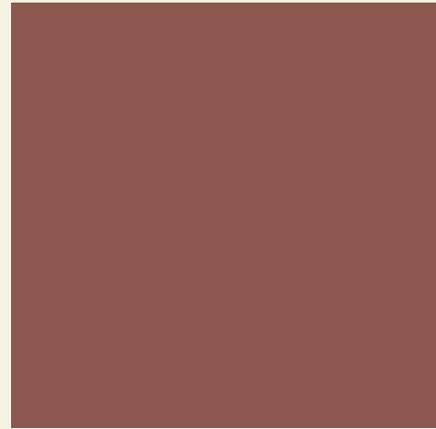
# ICCPP-STATISTICS

- Linear Regression

## Vishal Lohchab

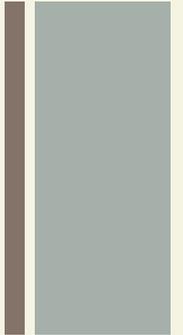
*Scientific Assistant of  
Prof. Dr. Hans-Werner Gessmann  
Director ICCPP International*





Francis Galton  
(1822-1911)  
Linear Regression

# + Definition



- Linear regression is the most widely used statistical technique; it is a way to model a relationship between two sets of variables.
- Linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.

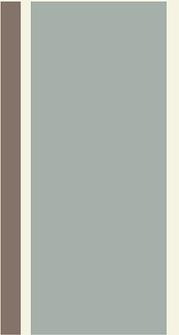
# + Formula

$$y = \alpha + \beta x$$

- $\beta$  = slope
- $\alpha$  = y-intercept
- $Y$  = y- coordinate
- $x$  = x-coordinate

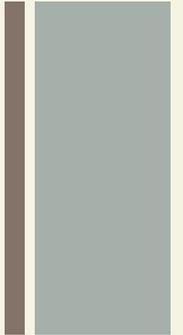
# + Use

- Simple linear regression is used to estimate the relationship between two quantitative variables. You can use simple linear regression when you want to know:  
How strong the relationship is between two variables (e.g. the relationship between rainfall and soil erosion).





# Assumptions

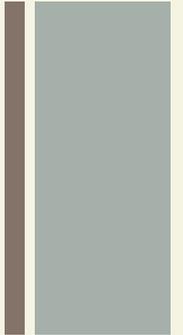


Simple linear regression is a parametric test, meaning that it makes certain assumptions about the data.

These assumptions are:



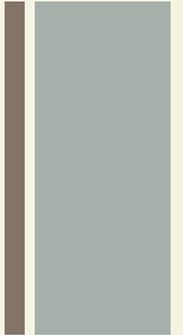
# Assumptions



- Homogeneity of variance (homoscedasticity): the size of the error in our prediction doesn't change significantly across the values of the independent variable.



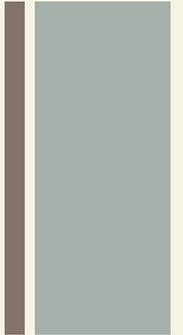
# Assumptions



- Independence of observations: the observations in the dataset were collected using statistically valid sampling methods, and there are no hidden relationships among observations.



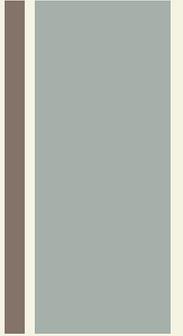
# Assumptions



- Normality: The data follows a normal distribution.
- The relationship between the independent and dependent variable is linear: the line of best fit through the data points is a straight line (rather than a curve or some sort of grouping factor).



# How to Find a Linear Regression Equation

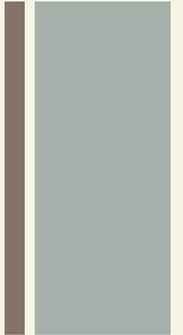


## Step 1

Make a chart of your data, filling in the columns in the same way as you would fill in the chart if you were finding the Pearson's Correlation Coefficient.



# How to Find a Linear Regression Equation: Steps

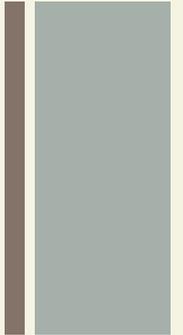


## Step 1.1

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	$x^2$	$y^2$
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
$\Sigma$	247	486	20485	11409	40022



# How to Find a Linear Regression Equation: Steps



## Step 1.2

From the table

$$\sum x = 247$$

$$\sum y = 486$$

$$\sum xy = 20485$$

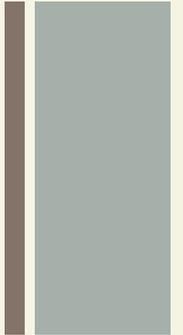
$$\sum x^2 = 11409$$

$$\sum y^2 = 40022$$

$n$  is the sample size (in our case 6)



# How to Find a Linear Regression Equation: Steps



## Step 2

Use the following equations to find a and b.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

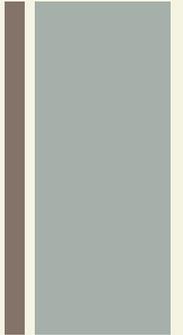
$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = 65.1416$$

$$b = .385225$$

+

# How to Find a Linear Regression Equation: Steps



## Step 2.1

Find a

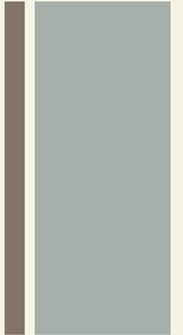
$$((486 \times 11,409) - ((247 \times 20,485)) / 6 (11,409) - 247^2) =$$

$$484979 / 7445$$

$$=65.14$$



# How to Find a Linear Regression Equation: Steps



## Step 2.2

Find b

$$\frac{(6(20,485) - (247 \times 486))}{(6(11,409) - 247^2)} = \frac{(122,910 - 120,042)}{68,454 - 247^2} =$$

$$2,868 / 7,445 = \mathbf{.385225}$$

# + Solution Steps wise

## Step 3

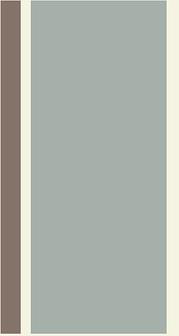
Insert the values into the equation.

$$y' = a + bx$$

$$y' = 65.14 + .385225x$$

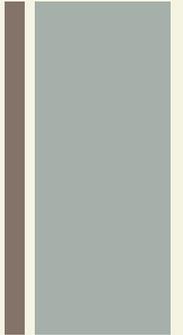
# + Solution Steps wise

- That's how to find a linear regression equation by hand
- \* Note that this example has a low correlation coefficient, and therefore wouldn't be too good at predicting anything.





# References



Edwards, A. L. *An Introduction to Linear Regression and Correlation*. San Francisco, CA: W. H. Freeman, 1976.

Edwards, A. L. *Multiple Regression and the Analysis of Variance and Covariance*. San Francisco, CA: W. H. Freeman, 1979.

Stephanie Glen. "Welcome to Statistics How To!"  
From StatisticsHowTo.com: Elementary Statistics for the rest of us!  
<https://www.statisticshowto.com/>