



Essential Math for Data Analysis Using Excel Online

Module 1, Lab 1: Data Set Anatomy

Learning Objectives

- Become familiar with the layout of a data set.
- Be able to find and identify variables and individuals.

Description

Learners will open up a data set with three variables. They will be asked to identify columns as variables and rows as individuals. They will be asked to perform a few tasks consistent with this, such as “what is the age of Person 15?”

Data set

Mod1Lab1.csv

Overview

In this lab, we'll get familiar with the layout of a data set. Interpreting data is an important first step before manipulating it, so there's no math involved just yet.

Exercise 1: Reading Data

1. Open the data set in Excel. Here's what our tidy data set looks like, with a few different variables:

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N

- Identify the variables. In this case, there are three different vertical **columns**: ID, age, and entry level. Those are the variables because those values vary from person to person.
- Identify the observations. Since each horizontal **row** corresponds to a single entry from each of the ID, age, and entry level columns, each row must represent an individual person. In other words, Person 1 is 45 years old and has selected N (“No”) for entry level, which means that person is not an entry-level employee.

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N

4. Now you can answer simple questions about the data set. For example, how old is Person 15? To figure it out, hunt down 15 in the ID column, then move along Person 15's row to the age variable. Heads up: You *don't* want to look at row 15 in the Excel doc. Person 15 is actually in row 16 because the ID variable tells us the person's number, not the row number.

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N

So Person 15 is 32 years old.

- Is Person 7 an entry-level employee or not? Head to the row for Person 7 and check the “entrylevel” column.

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N

Person 7 is not entry-level.

- How many entry-level employees are in the data set? To find out, count the number of Y's in the "entrylevel" column.

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N
22			

There are 3 entry-level employees in this data set.