Fundamentals of Statistics

Intellectus

Introduction

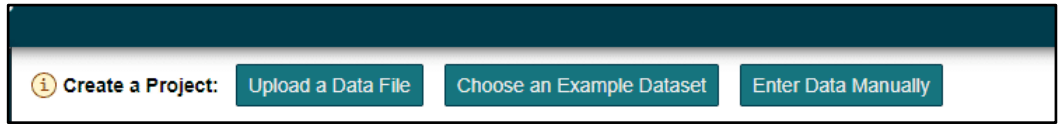
Opening Example

Choosing an example dataset

Throughout this module, you will be asked to open and run analyses on various example datasets supplied by Intellectus. If you are unsure of how to find an example dataset, follow the steps listed below.

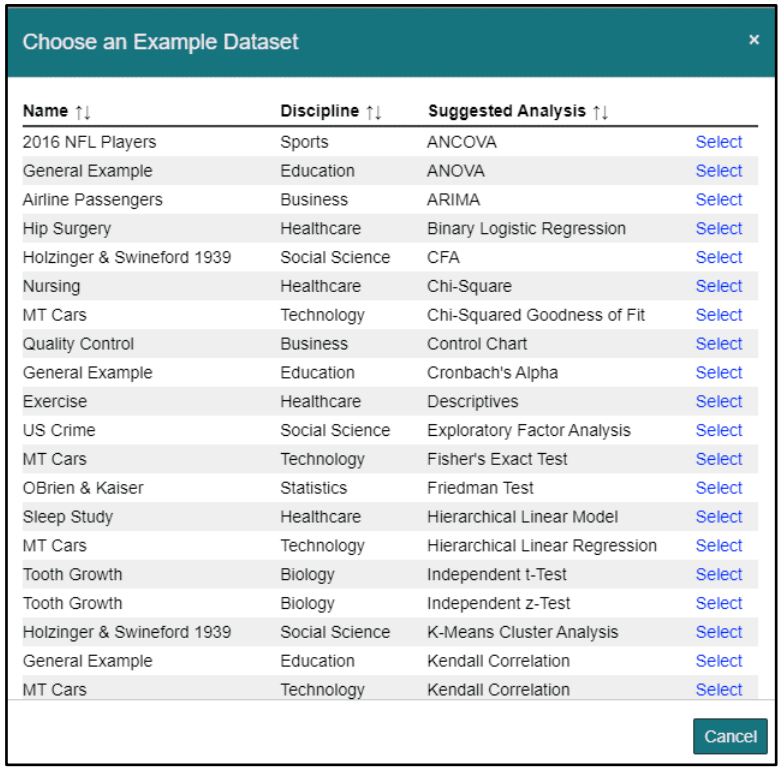
Step 1

Select "Choose and Example Dataset"



Step 2

Selecting the dataset

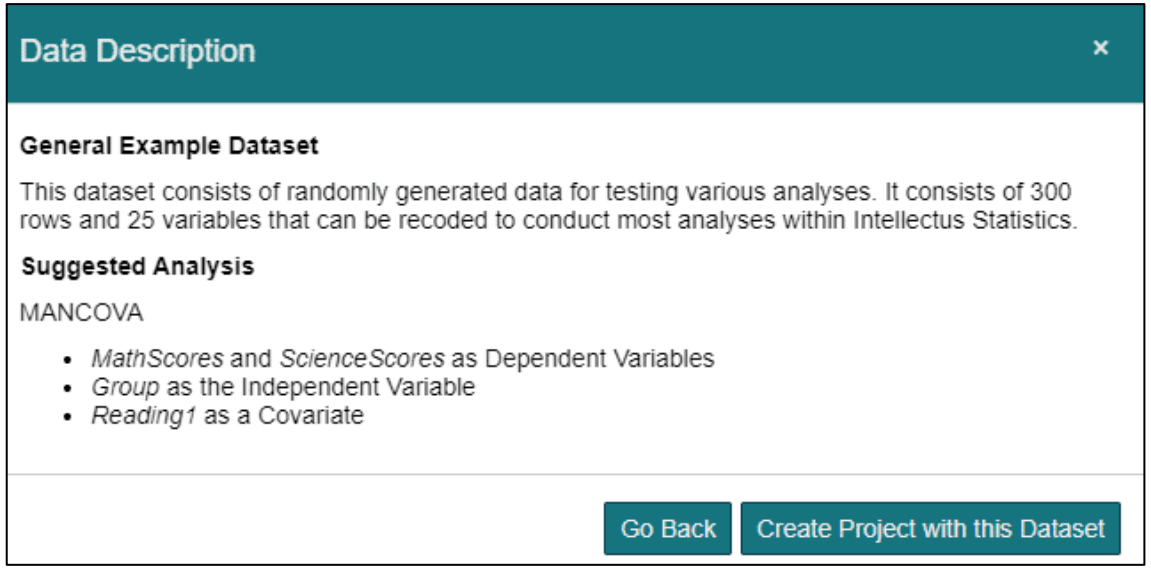


There are three columns: the name of the dataset, discipline, and suggested analysis. For the most part, we will be looking at the suggested analysis column in order to determine which dataset to open.

Opening Example

Step 3

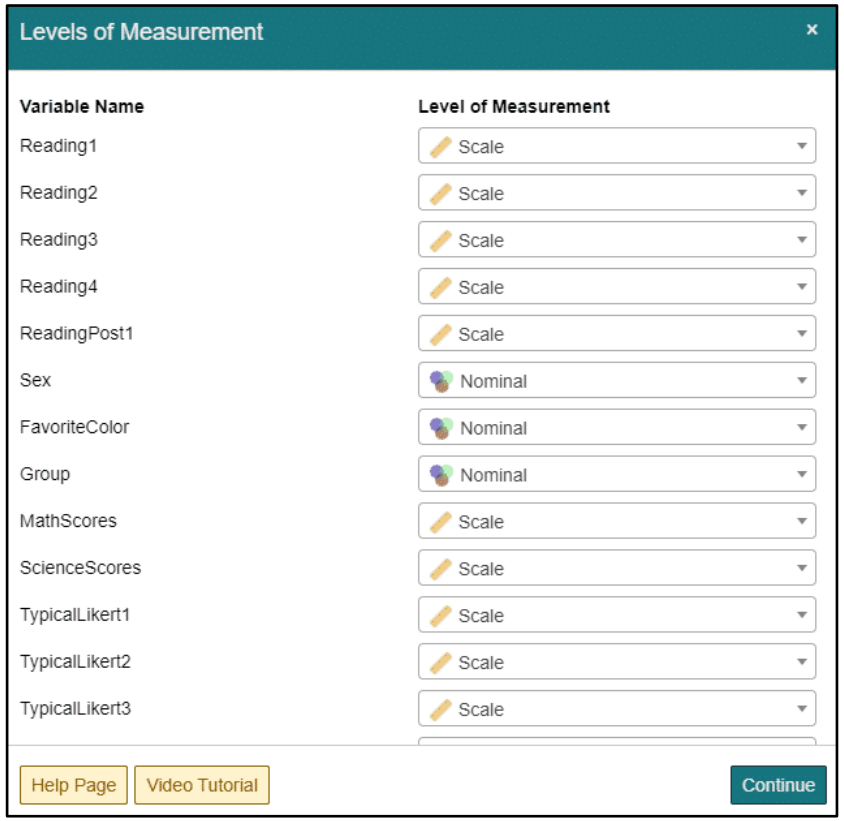
Details of the dataset



After opening the dataset, you will be presented with a description of the dataset and details of the suggested analysis. Select "Create a Project with this Dataset."

Step 4

Levels of measurement



Here you can manage the levels of measurement for each variable. For the examples presented in this module, these settings should be left alone.

Overview

What is Statistics?

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of numerical data.

Data is present nearly everywhere in our daily lives.

- Daily commute time
- Student test scores
- The heights of a given population
- The number of steps taken by individuals
- The monetary values of all the cars on the street

Data Collection and Measurement

Data Collection and Measurement

Why do we measure and analyze data?

Consider the following scenario:

Imagine you are a scientist attempting to determine if milk is good or bad for humans. To determine this, you give it to your child Mike, who grows up big and strong with healthy bones. Milk must be good for you - just look how Mike turned out. But then you hear about your friend's child Johnny, who gets stomach aches and feels ill every time he drinks milk. Now it seems milk might not be good for humans after all. You don't really have a way of seeing exactly what the milk is doing in each child, so how can you make a definitive statement for all humans?

The answer is statistics.

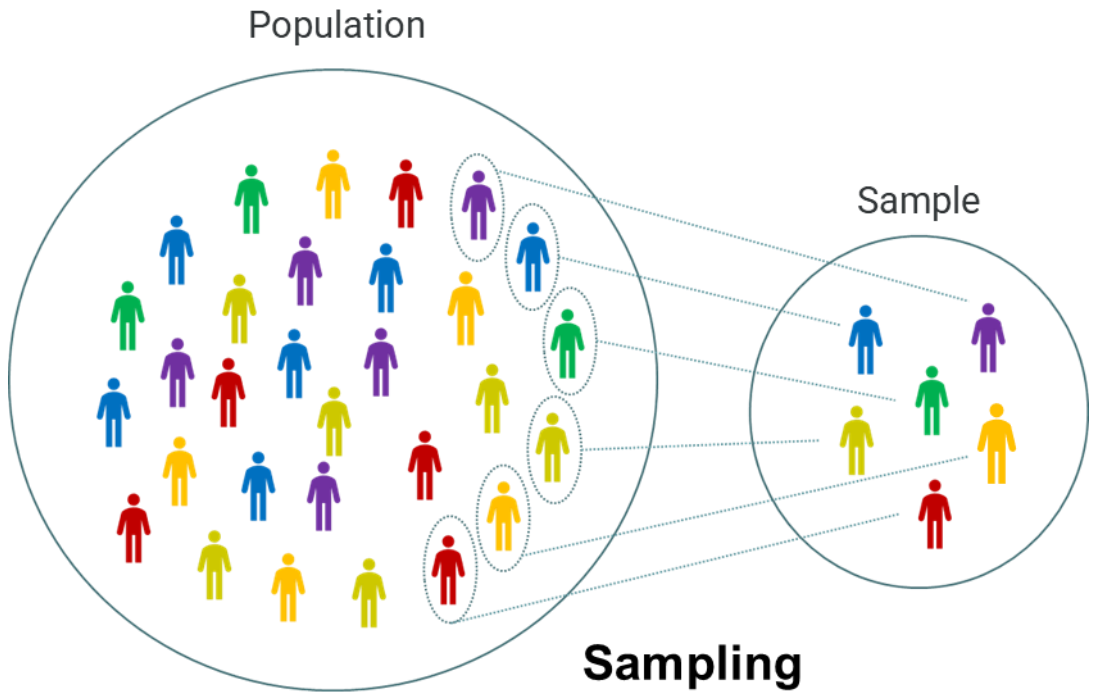
Milk will not affect everyone the same, so the key to determining if it should be considered healthy is to **measure** the effects across a large number of humans, and then **analyze** these results to see if there are any trends or differences.

What to Measure

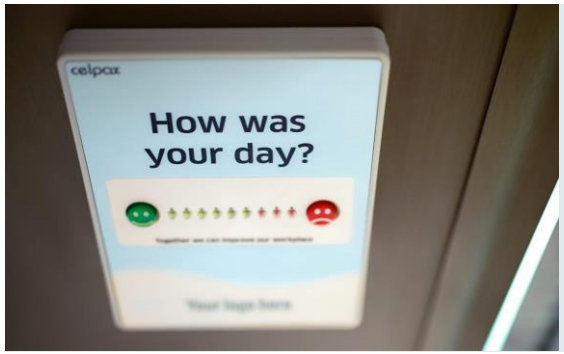
It is impossible to measure everything.

Being the determined scientist that you are, you continue with your research on milk. You figure that if you can record the effects of milk on everyone, you will know for sure the answer to whether or not milk is healthy.

But this is not possible. You only have so much time and only know so many people. So instead, you observe the effects of milk on all the people you are able to (your friends and family, and anyone else willing to let you watch them drink milk). You have just engaged in something called **sampling**.



Sampling



Definition:
Sampling is the act of measuring and recording a pre-determined number of observations from a population or process.

In our milk example, the sample is anyone you were able to observe the effects of milk on. Below are some terms and facts associated with sampling.

- A single recorded measurement is called an **observation**.
- The number of observations you have is called the **sample size**.
- Generally, the larger the sample size, the more that will be revealed about the population as a whole. A good way to think about this is to imagine a painting. Do you know more about what the painting depicts if you can only see a tiny corner, or if you can see all but the edges?

Your measurement of the effect milk on Mike is an observation. If you manage to observe the effects of milk on 99 other people, you have a sample size of 100. The more people you can observe the effects of milk on, the more you will learn about whether it should be considered healthy.

Review

Summary

Term	Definition
Sampling	The act of measuring and recording a predetermined number of observations from a population or process.
Observation	Single recorded measurement.
Sample size	The number of observations.