# Slide 1 – Title Slide

Hello and welcome to Week 5, Part 1 of EGM101: Introduction to Quantitative Skills. In this lesson, we're going to start learning about quantitative skills, and how we can apply them in Geography and Environmental Sciences.

# Slide 2 – Week 5 Outline

In the remaining parts of this week's lecture, we'll talk about data: how we can present it so that other people can make sense of it and use it, as well as start to talk about how we collect it. We'll also discuss the concepts of frequency and descriptive statistics, before finally wrapping things up by talking about data distributions.

# Slide 3 – Aims of Quantitative Skills Section

Before we jump right in though, we'll talk about the aims of this section of the module. EGM101 is designed to provide you with a wide range of skills that you will use and build on throughout the rest of your course. This includes mapping, which you have spent the last three weeks working on, quantitative skills, and qualitative skills, which you will cover in the final four weeks of the module.

One important thing to remember is that the goal of this part of the module is not to turn you into a statistician or a mathematician overnight, or even at all. Instead, the aim is to *introduce* you to some basic statistical methods for data analysis. Another aim is to introduce you to research data management, to help you keep your data organized and easier to keep track of.

And finally, through the practicals in this module, we will introduce you to selected software for working with data (which also goes by the sinister-sounding 'data manipulation'), analyzing data, and finally, presenting data.

# Slide 4 – Why Statistics?

So, why statistics? What's so special about this branch of mathematics that we feel the need to start off your very first semester in the course learning about it? Well, as we will see, numerical data are literally everywhere, and it keeps getting worse by the day.

Not only that, but the real world is noisy, or messy. One reason that the real world is "noisy" is because of natural variation. For example, let's say we wanted to keep track of the amount of snow that falls in a given place over a period of time. We might see that in one year, we get a lot of snow, while in other years, we get very little, or even no snow. How, then, would we describe the amount of snow that our location gets? That's something that statistics can help us with.

Another reason that the real world is noisy is because of uncertainty or error in our observations – maybe the way that we measure snowfall isn't very reliable, or maybe we put our instrument in a spot that is a bit sheltered. How to deal with inaccurate observations is something that statistics can help us with, too.

And, sadly, someone might also be trying to deceive us – by making up fake data, or maybe by cheating in a video game. By making use of statistical methods, we might be able to spot the lies or cheating.

All of this is to say that statistics, and quantitative skills more generally, help us to work with our data, cut through some of that noise, and make sense of what we're looking at. But maybe not too much.

## Slide 5 – Some Definitions

Before we go too much further, I think it's a good idea to introduce some definitions, starting with statistics. As I alluded to on the previous slide, statistics is the science of collecting, analyzing, and presenting data – it's how we work with the data and observations that we collect. We can break "statistics" down into two categories. The first, descriptive statistics, is how we organize and summarize our data, and it's what we'll spend most of the rest of this week covering.

The second, inferential statistics, is how we can draw conclusions from "good" data – and later on, we'll discuss what we mean by "good" data.

The next definition we'll introduce is "population" – this is the entire set or group of what it is that we're studying. So, if we're studying salmon, we can think of the population as all of the salmon in the lake, river, or ocean that we're studying. Connected to this is a variable, which is any characteristic that has different values for each member of the population. Sticking with our salmon example, this could be something like length, weight, age, sex, color, or any other number of attributes. In humans, this could also be things like voting intention or opinion.

Now, the problem is, there are usually a lot of members of a population, and we normally can't observe or measure all of them. That means that we need to take a sample, or some portion of the population, that we can observe.

Finally, we have a parameter, which is a typically numeric characteristic or property of the population. This could be something like the average weight of all of the salmon in the lake, for example. Because we don't normally work with the entire population, though, we instead work with a statistic, which is a number that represents a property of the sample. Because the sample is a representation of the population, the statistic is an approximation of the parameter. We'll talk more about how samples relate to a population, and how a statistic relates to a parameter, as we continue through this part of the module – for now, it's enough to remember that a statistic belongs to a sample, and a parameter belongs to the population.

## Slide 6 – Types of Variables (Data)

On the previous slide, we defined a variable as a characteristic or attribute that has a different value for each member of the population. You can also think of variables as the data we work with – for example, in the first practical, we worked with a dataset with multiple temperature variables – the monthly maximum and minimum monthly temperatures recorded at the Armagh Observatory since 1853.

We can split variables, or data, into two broad categories. The first of these categories is qualitative, or categorical, data. We can break qualitative data into two further categories: nominal data, and ordinal data. Qualitative data are typically data that have non-numeric values, such as categories or descriptions of attributes.

The second broad category of data is quantitative, or numerical data. Quantitative data are always numbers, and we can further sub-divide this into discrete data, and continuous data. On the next two slides, I'll provide some examples of these different types of data to help illustrate the differences between them.

## Slide 7 – Qualitative Data

The first type of qualitative data we'll discuss is nominal data, which we use for categories that don't have a particular order. Some different examples of nominal variables are eye color – for example, we might have blue eyes, or also blue eyes, or even red-orange eyes. Another example of nominal data might be blood type, or type of pet.

These distinctions – blue, red-orange, green, or type O positive, O negative, and so on – don't have a particular order or value attached to them – this is what we mean when we say "unordered". If we do have a particular order for qualitative data, then what we have is an ordinal variable. For example, race or election results, where an athlete or political candidate finishes first, second, third, and so on, would be ordinal data – there's a specific order to the place they finish. Opinions, where a survey respondent might agree or disagree with a specific question or statement, is another example of ordinal data, as are school marks or rings of power – here, we have 16 lesser rings for the Dwarf-lords and Mortal Men, 3 rings for the Elven-kings under the sky, and one ring to rule them all.

## Slide 8 – Quantitative Data

That brings us to quantitative data – and again, these are always going to be numbers. Discrete variables are data that can only take on certain values. Things like population or fish counts are examples of discrete variables, because we can only ever count whole numbers of people or fish. Most of the time, if we're dealing with counting people or animals or objects, we're dealing with discrete variables.

The other main type of quantitative variable we'll work with are continuous variables – these are variables that can take on any real-valued number. By "real-valued", I mean numbers that have a decimal value – -1, or 42.0, or pi, are all examples of real-valued numbers. Some examples of continuous variables we might see are things like fish length, or temperature.

## Slide 9 – Summary

In this lesson, we discussed the idea that the real world is messy, and we can't possibly measure everything.

Fortunately, as we'll see through the rest of this part of the module, statistics can help us make sense of things, by enabling us to measure only a portion of everything, and figure out how to make some sense of the mess.

We also saw how we can divide data or variables into different categories, based on what it is that we're working with. In this part of the module, we're mostly going to concern ourselves with quantitative, or numerical, data. It's kind of in the name, after all.

## Slide 10 – Additional resources

You can read more about the topics we've discussed here in the textbooks – in Illowsky and Dean, Chapters 1.1 and 1.2, Caswell Chapter 4.1, and Weiss, Chapters 1.1 and 2.1.

I've also included some links to a few different youtube videos here – the first is a lecture by the mathematician Hannah Fry, discussing some of the many different applications of math (or maths, if you prefer). The second is a slightly longer lecture by the statistician Jennifer Rogers, discussing many of the topics that we'll touch on in this part of the module. The next video is a TED talk given by the data journalist Alan Smith, talking about why you should love statistics, and the final video is by the comedian and mathematician Matt Parker, discussing an alleged cheating scandal in Minecraft. We will talk a bit more about some of the concepts introduced in that video in Week 7 when we discuss probability, so don't worry too much now if it feels a bit over your head.

That's all for this lesson – I hope you found it interesting, and you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!