Reading 7: Statistical Concepts and Market Returns

LOS 7a: distinguish between descriptive statistics and inferential statistics, between a population and a sample, and among the types of measurement scales

Statistics refers to a set of concepts, rules, and procedures that help us to interpret data and make informed decisions regarding issues affecting our lives and humanity in general. Data is basically facts or observations that result from an investigation. Statistics is a wide field of study that has relevance across most areas including the economy, academic as well as industrial research. There are two main branches of statistics — descriptive and inferential statistics.

Descriptive Statistics

Descriptive statistics provide simple summaries about data. They represent complex data in a manageable form and do not go as far as generalizing the data. A good example is the average height of men in a country. Measures of descriptive statistics include the mean, variance, kurtosis, and skewness.

Inferential statistics

This entails using the statistical sample data to draw **valid conclusions** concerning the entire population. Under inferential statistics, we have probability distributions, hypothesis testing, correlation/regression analysis, and probability distributions.

Population vs. Sample

A population refers to the summation of all the elements of interest to the researcher. These could be the number of people in a country, the number of animals in a zoo or the total number of cars in a given county.

A sample is just a set of elements that represent the population as a whole. It's very difficult to collect data from each and every element in the population. This calls for use of a sample comprising a small number of representative elements. By analyzing sample data, we are able to make conclusions about the entire population. Thus, sampling makes the analysis of a large population manageable and help researchers save time.

Types of Measurement Scales in Statistics

We use measurement scales to categorize variables. There are four main categories:

1. **Nominal scales**

Nominal scales are used to label variables but have no inherent numerical significance. A good example would be gender representation, e.g., 1 to represent 'male' and 2 to represent 'female'.

2. **Ordinal scales**

Ordinal scales are used to represent ordered categories where each category has an ordered relationship to all the other categories. Note that it's the order of categories that is important but not the differences between them, which cannot be quantified. A good example is a rating scale starting from 1 to 3, where 1 represents ok, 2 for good and 3 for excellent.

3. **Interval scale**

With interval scales, we know both the order and the exact differences between the values. A good example of this type of scale is time because we can establish the exact increments or decrements. An even clearer example is temperature measurement in Celsius – the difference between, say, 60 and 80 degrees is a measurable 20 degrees.

4. **Ratio scales**

Ratio scales are complete in the sense that they tell us about the order, the exact differences between values and also have a true zero. Thus, we can calculate ratios. Height and weight are good examples of ratio variables. Consequently, ratio scales pave the way for measures of central

tendency like the mode and the mean, as well as measures of dispersion, e.g., standard deviation.

Reading 7 LOS 7a

Distinguish between descriptive statistics and inferential statistics, between a population and a sample, and among the types of measurement scales.

Quantitative Methods – Learning Sessions

LOS 7b: define a parameter, a sample statistic, and a frequency distribution

A parameter refers to a measure that is used to describe a characteristic of the **population.** It's a numerical quantity that describes a given aspect of the population as a whole. Candidates should note that we are referring to the population as a whole, not a sample. There are many parameters that could be used to measure different characteristics of the population but just a few of them are used regularly by investment analysts. These include the mean and the standard deviation of investment returns. (Examiners concentrate questions around these parameters.)

A statistic, on the other hand, is a measure that describes a characteristic of a **sample.** This could be the average value or the sample standard deviation of the sampled items.

Note: Researchers use sample statistics to estimate the unknown population parameters. For instance, we use the sample mean to estimate the actual population mean.

Frequency Distribution

A frequency distribution refers to the presentation of statistical data in a tabular format to simplify the analysis. The data is subdivided into groups or intervals. The standard procedure for constructing a frequency distribution involves the following steps:

1. Determine the number of classes you wish to have. 5 – 20 is always a good number.

2. Determine the interval size. To do this, you should be guided by the range and the number of classes. The range is the difference between the smallest and the largest observations. In the case of a fractional result, you should take the next higher whole number as the size of the interval.

3. Determine the starting point – it could be the lower limit of the lowest observation or a convenient value just below the lower limit.

4. Add the class interval to the starting point so as to get the second lower limit and repeat this process.

5. List the lower limits in a vertical column alongside the upper-class limits.

6. You can now complete the table by counting the number of observations that fall under each class.

Points to note: Classes should be mutually exclusive and have similar widths. Also, you must tabulate all class intervals even if they have zero observations. The sum of the frequencies should be equal to the number of observations. Tally bars offer a convenient tool for visual presentation of the number of observations in each class.

Question

You have been given the following data that show the percentage returns offered by certain classes of investment in a certain year. Use the data to construct a frequency distribution table.

-10%	2%	32%	-28%	25%
-25.60%	4%	11%	-14%	15%
23%	13%	6%	-2.70%	8%
12%	28%	17.50%	5.80%	20%
4.60%	17%	-3.90%	22.40%	15%

Solution

We have 25 observations in total. The range is 60% (- 28% to 32%). If we were to choose an interval of just 1%, we would end up with 60 intervals which would be too many. So, we can use an interval of 10%. The lowest return intervals will be -30% ≤ R_t < -20% and the highest one will be 30% ≤ R_t < 40%.

Interval	Tally	Frequency
$-30\% \leq R_t < -20\%$	II	2
$-20\% \leq R_t < -10\%$	I	1
$-10\% \leq R_t < 0\%$	III	3
$0\% \leq R_t < 10\%$	IIIIII	6
$10\% \leq R_t < 20\%$	IIIIIII	7
$20\% \leq R_t < 30\%$	IIIII	5
$30\% \leq R_t < 40\%$	I	1
Total		25

Reading 7 LOS 7b

Define a parameter, a sample statistic, and a frequency distribution

---

LOS 7c: calculate and interpret relative frequencies and cumulative relative frequencies, given a frequency distribution

Relative frequency refers to the percentage of observations falling within a given class. It reveals the popularity of certain classes of data based on a sample. In other words, relative frequency tells us the number of times an event occurs **relative** to the **total** number of events. Using our earlier example when we introduced the frequency distribution table, we could come up with the relative frequency for each interval using the formula:

Relative frequency = Absolute frequency / Total frequency

Total frequency is just the total number of observations.

Example 1: Relative Frequency

Interval	Tally	Frequency	Relative Frequency
$-30\% \leq R_t \leq -20\%$	II	2	= 2/25 = 8%
$-20\% \leq R_t \leq -10\%$	I	1	= 1/25 = 4%
$-10\% \leq R_t \leq 0\%$	III	3	= 3/25 = 12%
$0\% \leq R_t \leq 10\%$	IIIIII	6	= 6/25 = 24%
$10\% \leq R_t \leq 20\%$	IIIIIII	7	= 7/25 = 28%
$20\% \leq R_t \leq 30\%$	IIIII	5	= 5/25 = 20%
$30\% \leq R_t \leq 40\%$	I	1	= 1/25 = 4%
Total		25	= 25/25 = 100%

Furthermore, we could come up with cumulative frequencies. The cumulative absolute frequency is the sum of the absolute frequencies up to and including the given interval. The cumulative relative frequency similarly sums up the relative frequencies up to and including the given relative frequency.

Example 2: Cumulative Frequencies

Interval	Tally	Frequency	Relative Frequency	Cumulative absolute frequency	Cumulative relative frequency
$-30\% \le R_t \le -20\%$	II	2	= 2/25 = 8%	2	8%
$-20\% \le R_t \le -10\%$	I	1	= 1/25 = 4%	3	12%
$-10\% \le R_t \le 0\%$	III	3	= 3/25 = 12%	6	24%
$0\% \le R_t \le 10\%$	IIIIII	6	= 6/25 = 24%	12	48%
$10\% \le R_t \le 20\%$	IIIIIII	7	= 7/25 = 28%	19	76%
$20\% \le R_t \le 30\%$	IIIII	5	= 20/25 = 8%	24	96%
$30\% \le R_t \le 40\%$	I	1	= 1/25 = 4%	25	100%
Total		25	= 25/25 = 100%		

Reading 7 LOS 7c

Calculate and interpret relative frequencies and cumulative relative frequencies, given a frequency distribution

---

Quantitative Methods – Learning Sessions