Statistics for Data Science - 1

FAQ

Week 1

1. Why are jersey number, mobile number, and pincode nominal scales of measurement?
   Answer: A categorical variable is said to have a nominal scale of measurement if there is no order in the categories. Generally, names and labels do not have an order. Hence, they are said to have nominal scale of measurement.

   For instance, consider the jersey numbers of cricket players. There is no concept of ordering for these jersey numbers. As you can see, Dhoni's jersey number 7 is in no way greater than or lesser than Gambhir's jersey number 5. Further, it is meaningless to perform mathematical operations on any two jersey numbers. A player with jersey number 12 is not the sum of two players whose jersey numbers add up to 12 (say jersey numbers 5 and 7).

   Similarly, even though mobile numbers and pin codes are numbers, it is meaningless to associate an order to these numbers. There is no order in pin codes 100002, 500001, 500002 i.e., pin code 100002 is neither greater nor lesser than 500001 or 500002. However, if the numbers 100002, 500001 and 500002 represent money in rupees, then the ordering $100002 < 500001 < 500002$ makes sense. In this case, we have an order. The same explanation applies to mobile numbers too.

2. How does the conversion of interval scale of measurement to ratio scale of measurement by subtraction work?
   Answer: The conversion of interval scale to ratio scale by a subtraction operation is an application of what Prof. Usha Mohan has taught in her lectures. Variables with an interval scale of measurement have a fixed unit of measure and do not have an absolute zero (as in the case of degree Celsius). So, multiplication and division operations can not be performed on these variables.
   Variables with a ratio scale of measurement have a fixed unit of measure and absolute zero (temperature in Kelvin). So division and multiplication operations can be performed.

   The only difference between ratio and interval scales is with respect to absolute zero. If a variable with an interval scale of measurement has absolute zero, it will have a ratio scale of measurement. Absolute zero exists when we subtract two numerical values with an interval scale of measurement. For example, in Restaurant 1 and Restaurant 2, the rating given by users is recorded. The rating given by a user should be an integer from 1 to 5. Since the ratings given by users have a fixed measure and no absolute zero, it follows that the rating is an interval scale.
   However, if we are interested in the difference (absolute value) between ratings given by the same user, then the new variable can take values from 0 to 4. This variable has

absolute zero. So, it has a ratio scale of measurement. Table 1 gives an example of ratings.

S.No.	User	Restaurant 1	Restaurant 2	Difference in rating
1.	User 1	4	1	3
2.	User 2	3	5	2
3.	User 3	5	5	0
4.	User 4	3	4	1
5.	User 5	5	3	2

Table 1: Restaurant ratings

3. Why is it that the interval scale of measurement can take negative values but not the ratio scale of measurement?
   Answer: Interval scales do not have absolute zero and can take negative values, whereas ratio scales do not have negative values.
   In the Kelvin scale, temperature cannot be less than zero and hence cannot be negative. In other words, zero Kelvin is the minimum value of temperature possible and hence temperature in Kelvin scale has ratio scale of measurement.
   In the case of temperature measured in Celsius scale, the values can be negative. It can take up values as low as -273 degree Celsius.

4. Give an example of interval scale other than the temperature example.
   Answer: Any numerical variable will have an interval scale of measurement if it has all the properties of ordinal scale and the difference between the values is expressed as a fixed unit of measure.
   Consider the example of GPA (Grade Point Average) received by students based on their performance in academics. The GPA is in the range [4,10]. We see that although there is ordering in GPA, it doesn't have an absolute zero. We can not say that 8.0 GPA is twice as good as 4.0 GPA. So, Grade Point Average (GPA) has an interval scale of measurement.

5. What is the difference between ordinal scale of measurement and nominal scale of measurement?
   Answer: For a categorical variable to have an ordinal scale, irrespective of frequencies there should be order (rank) in the categories.
   The feedback given by customers in a fruit shop is either *Good*, *Bad* or *Average*. On a Sunday, the ratings given by a set of 45 customers are as follows:

   - Good - By 30 customers
   - Bad - By 15 customers
   - Average - By 5 customers

The descending order based on frequencies is Good > Bad > Average. But we know that the original order of categories is Good > Average > Bad. The shop owner considers Good as highest and Bad as lowest, with Average being in between Good and Bad. Hence, ratings by the customer is an ordinal scale of measurement. We now give an example of nominal scale of measurement. Table 2 gives the quantity of each fruit sold in a fruit shop on a particular day.

S.No.	Item sold	Quantity (kg)
1.	Kiwi	5
2.	Apple	14
3.	Orange	20

Table 2: Fruit shop data

Even though more oranges were sold than either apples or kiwis, we know that orange is neither greater than nor smaller than apple or kiwi. The shop owner considers every fruit the same i.e., there is no order among the fruits. Hence, item sold is a nominal scale of measurement.

6. What is the difference between inferential statistics and descriptive statistics?
   Answer: Descriptive statistics deals with describing and summarizing the given data using certain parameters while inferential statistics deals with arriving at conclusions from the data. For example, if you want to describe the performance of students in an exam conducted yesterday, then it comes under descriptive statistics. On the other hand, if you want to infer the performance of the students in the exam to be conducted tomorrow, using the data of the exam conducted yesterday, then it comes under inferential statistics.

7. What is the difference between cross-sectional and time series data?
   Answer: The temperature in all major cities of India recorded at a particular instant of time is cross-sectional data. On the other hand, the temperature in Delhi recorded throughout the day is time series data. That is, if the values in the dataset vary with respect to space, with the time being constant, it is cross-sectional data. If the data varies with respect to time for a particular entity in space, it is time series data.

8. What is the difference between ordinal scale of measurement and interval scale of measurement?
   Answer: In the case of ordinal scale of measurement, the difference between two consecutive pairs of values need not be the same whereas in the case of interval scale of measurement, the difference between two consecutive values is a fixed unit of measure. In the example given in Table 1, the possible ratings are integers from 1 to 5. The ordered values 1, 2, 3, 4, 5 have a difference of 1 between any two consecutive values. Therefore, the ratings dataset has an interval scale of measurement.
   If the options for ratings are Good, Average, and Bad, then it has an ordinal scale

of measurement since we do not know for sure that the difference between Good and Average is same as the difference between Average and Bad.