### **Statistics Fundamentals**

1. **Variables in Statistics**

**Mission 1: Variables in Statistics**

Previously, we discussed the details around collecting data for our analysis. In this mission, we'll focus on understanding the structural parts of a data set, and how they're measured.

Whether a sample or a population, a data set is generally an attempt to describe correctly a relatively small part of the world. The data set we worked with in the previous mission describes basketball players and their performance in the season 2016-2017.

Other data sets might attempt to describe the stock market, patient symptoms, stars from galaxies other than ours, movie ratings, customer purchases, and all sorts of other things.

The things we want to describe usually have a myriad of properties. A human, for instance, besides the property of being a human, can also have properties like height, weight, age, name, hair color, gender, nationality, whether they're married or not, whether they have a job or not, etc.

In practice, we limit ourselves to the properties relevant to the questions we want to answer, and to the properties that we can actually measure. Let's consider three rows at random from the basketball data set we've previously worked with:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Name | Team | Pos | Height | Weight | BMI | Birth\_Place | Birthdate | Age | College | Experience | Games Played | MIN | FGM | FGA | FG% | 15:00 | 3PA | 3P% | FTM | FTA | FT% | OREB | DREB | REB | AST | STL | BLK | TO | PTS | DD2 | TD3 |
| 39 | Crystal Langhorne | SEA | F/C | 188 | 84.0 | 23.766410 | US | October 27, 1986 | 30 | Maryland | 10 | 30 | 848 | 160 | 240 | 66.7 | 1 | 2 | 50.0 | 49 | 68 | 72.1 | 35 | 140 | 175 | 46 | 16 | 11 | 50 | 370 | 2 | 0 |
| 52 | Érika de Souza | SAN | C | 196 | 86.0 | 22.386506 | BR | September 3, 1982 | 34 | Brazil | 13 | 30 | 579 | 65 | 112 | 58.0 | 0 | 0 | 0.0 | 29 | 32 | 90.6 | 58 | 74 | 132 | 35 | 18 | 7 | 37 | 159 | 0 | 0 |
| 102 | Nia Coffey | SAN | F | 185 | 77.0 | 22.498174 | US | May 21, 1995 | 22 | Northwestern | R | 25 | 203 | 16 | 59 | 27.1 | 0 | 4 | 0.0 | 16 | 22 | 72.7 | 16 | 30 | 46 | 6 | 5 | 6 | 14 | 48 | 0 | 0 |

Each row describes an individual having a series of properties: name, team, position on the field, height, etc. For most properties, the values vary from row to row. All players have a height, for example, but the height values vary from player to player.

The properties with varying values we call variables. The height property in our data set is an example of a variable. In fact, all the properties described in our data set are variables.

A row in our data set describes the actual values that each variable takes for a given individual.

Notice that this particular meaning of the "variable" concept is restricted to the domain of statistics. A variable in statistics is not the same as a [variable in programming](https://en.wikipedia.org/wiki/Variable_(computer_science)), or [other domains](https://en.wikipedia.org/wiki/Variable).

**Mission 2: Quantitative and Qualitative Variables**

Variables in statistics can describe either quantities, or qualities.

For instance, the Height variable in our data set describes *how tall* each player is. The Agevariable describes *how much* time has passed since each player was born. The MIN variable describes *how many* minutes each player played in the 2016-2017 WNBA season.

Generally, a variable that describes *how much there is of something* describes a quantity, and, for this reason, it's called a quantitative variable.

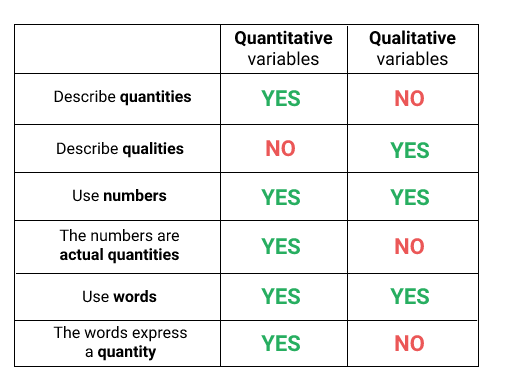
Usually, quantitative variables describe a quantity using real numbers, but there are also cases when words are used instead. Height, for example, can be described using real numbers, like in our data set, but it can also be described using labels like "tall" or "short".

A few variables in our data set clearly don't describe quantities. The Name variable, for instance, describes the name of each player. The Team variable describes what team each player belongs to. The College variable describes what college each player goes or went to.

The Name, Team, and College variables describe for each individual a *quality*, that is, a property that is not quantitative. Variables that describe qualities are called qualitative variablesor categorical variables. Generally, qualitative variables describe *what* or *how* something is.

Usually, qualitative variables describe qualities using words, but numbers can also be used. For instance, the number of a player's shirt or the number of a racing car are described using numbers. The numbers don't bear any quantitative meaning though, they are just *names*, not quantities.

In the diagram below we do a head-to-head comparison between qualitative and quantitative variables:



Instructions

We've selected a few variables from our data set. For each of the variables selected, indicate whether it's quantitative or qualitative.

* We've already created a dictionary named variables. Each variable name is given as dictionary key.
* If a variable is quantitative, then complete the value of the corresponding key with the string 'quantitative'. If the variable is qualitative, the use the string 'qualitative'.
* You can find useful documentation about each variable [here](https://www.basketball-reference.com/about/glossary.html) and [here](https://www.kaggle.com/jinxbe/wnba-player-stats-2017).

**Mission 3: Scales of Measurement**

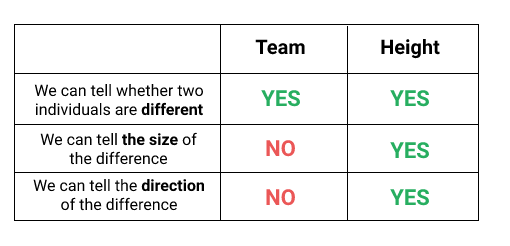
The amount of information a variable provides depends on its nature (whether it's quantitative or qualitative), and on the way it's measured.

For instance, if we analyze the Team variable for any two individuals:

* We can tell whether or not the two individuals are *different* from each other with respect to the team they play.
* But if there's a difference:
  + We can't tell *the size* of the difference.
  + We can't tell *the direction* of the difference - we can't say that team A is greater or less than team B.

On the other side, if we analyze the Height variable:

* We can tell whether or not two individuals are different.
* If there's a difference:
  + We can tell *the size* of the difference. If player A has 190 cm and player B has 192 cm, then the difference between the two is 2 cm.
  + We can tell *the direction* of the different from each perspective: player A has 2 cm *less* than player B, and player B has 2 cm *more* than player A.



The Team and Height variables provide different amounts of information because they have a different nature (one is qualitative, the other quantitative), and because they are *measured differently*.

The system of rules that define how each variable is measured is called scale of measurement or, less often, level of measurement.

In the next screens, we'll learn about a system of measurement made up of four different scales of measurement: nominal, ordinal, interval, and ratio. As we'll see, the characteristics of each scale pivot around three main questions:

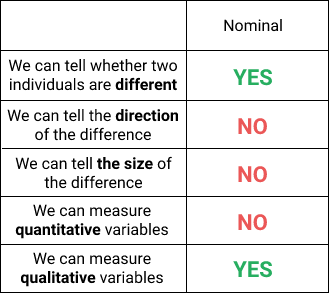
* Can we tell whether two individuals are different?
* Can we tell the direction of the difference?
* Can we tell the size of the difference?

**Mission 4: The Nominal Scale**

In the previous screen, we've discussed about the Team variable, and said that by examining its values we can tell whether two individuals are different or not, but we can't indicate the size and the direction of the difference.

The Team variable is an example of a variable measured on a nominal scale. For any variable measured on a nominal scale:

* We can tell whether two individuals are different or not (with respect to that variable).
* We can't say anything about the direction and the size of the difference.
* We know that it can only describe qualities.



When a qualitative variable is described with numbers, the principles of the nominal scale still hold. We can tell whether there's a difference or not between individuals, but we still can't say anything about the size and the direction of the difference.

If basketball player *A* has the number 5 on her shirt, and player B has 8, we can tell they're different with respect to shirt numbers, but it doesn't make any sense to subtract the two values and quantify the difference as a 3. Nor it makes sense to say that B is greater than A. The numbers on the shirts are just identifiers here, they don't quantify anything.

Instructions

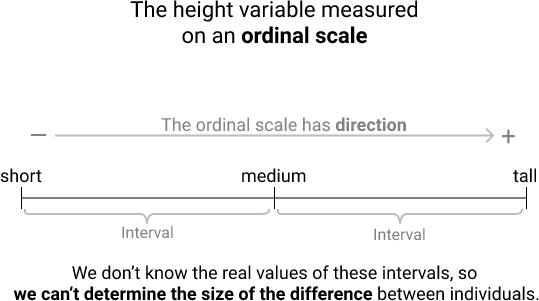
Inspect the data set, and find the variables measured on a nominal scale. In the code editor:

* Add the variables measured on a nominal scale to a list named nominal\_scale, and sort the elements in the list alphabetically (the sorting helps us with answer checking).
* Notice that we've added a new variable named Height\_labels. Instead of showing the height in centimeters, the new variable shows labels like "short", "medium", or "tall". By considering the principles that characterize the nominal scale, think whether the new Height\_labels variable should be included in your nominal\_scale list.

**Mission 5: The Ordinal Scale**

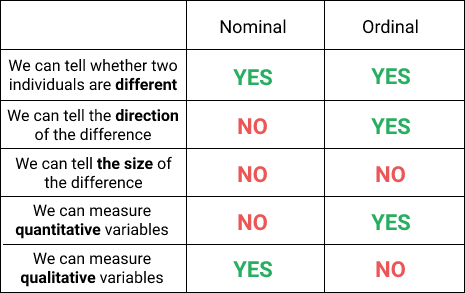
In our last exercise, we've seen that the new Height\_labels variable was showing labels like "short", "medium", or "tall". By examining the values of this new variable, we can tell whether two individuals are different or not. But, unlike in the case of a nominal scale, we can also tell the direction of the difference. Someone who is assigned the label "tall" has a bigger height than someone assigned the label "short".

However, we still can't determine the size of the difference. This is an example of a variable measured on an ordinal scale.



Generally, for any variable measured on an ordinal scale, we can tell whether individuals are different or not, we can also tell the direction of the difference, but we still can't determine the size of the difference.

Variables measured on an ordinal scale can only be quantitative. Quantitative variables, however, can be measured on other scales too, as we'll see next in this mission.



Common examples of variables measured on ordinal scales include ranks: ranks of athletes, of horses in a race, of people in various competitions, etc.

For example, let's say we only know that athlete A finished second in a marathon, and athlete B finished third in the same race. We can immediately tell their performance is different, we know that athlete A finished faster, but we don't know how much faster. The difference between the two could be half a second, 12 minutes, half an hour, etc.

Other common examples include measurements of subjective evaluations that are generally difficult or near to impossible to quantify with precision. For instance, when answering a survey about how much they like a new product, people may have to choose a label between "It's a disaster, I hate it", "I don't like it", "I like it a bit", "I really like it", "I simply love it".

The values of the variables measured on an ordinal scale can be both words and numbers. When the values are numbers, they are usually ranks. But we still can't use the numbers to compute the size of the difference. We can't say how much faster an athlete was than another by simply comparing their ranks.

Whether a variable is quantitative or qualitative is independent of the way the variable is measured. The Height variable, for instance, is quantitative no matter how we measure it. The fact that we use words like "short" or "tall" doesn't change its underlying nature. The Height variable still describes a magnitude, but in a different way.

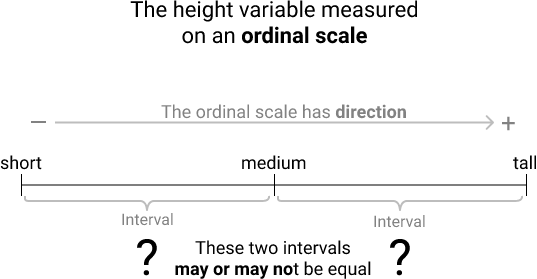
Instructions

Consider the following sentences, and evaluate their truth value. If the sentence is true, then assign True to the corresponding variable (programming variable) in the code editor, otherwise assign False. Make sure you assign boolean values as answers, not strings.

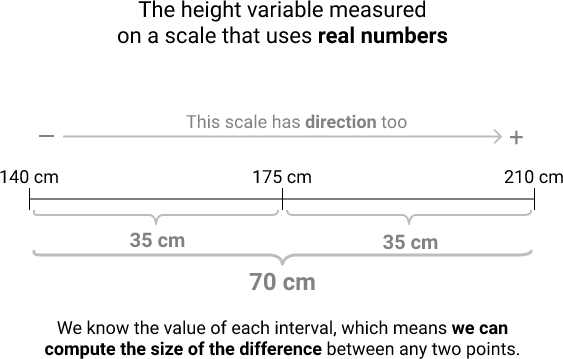
1. Using the Height\_labels variable only, we can tell whether player Kiah Stokes is taller than Riquna Williams. Assign your answer to a variable named question1.
2. We can measure the height difference between Kiah Stokes and Riquna Williams using the Height\_labels variable. Assign your answer to question2.
3. The Height\_labels and the College variables are both measured on an ordinal scale. Assign your answer to question3.
4. The Games Played variable is not measured on an ordinal scale. Assign your answer to question4.
5. The Experience variable is measured on an ordinal scale. Assign your answer to question5.
6. The Height\_labels variable is qualitative because it is measured using words. Assign your answer to question6.

**Mission 6: The Interval and Ratio Scales**

We've seen in the case of the Height variable that the values have direction when measured on an ordinal scale. The downside is that we don't know the size of each interval between values, and because of this we can't determine the size of the difference.



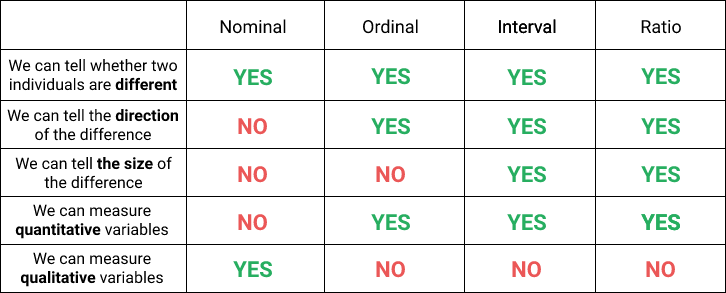
An alternative here is to measure the Height variable using [real numbers](https://en.wikipedia.org/wiki/Real_number), which will result in having well-defined intervals, which in turn will allow us to determine the *size* of the difference between any two values.



A variable measured on a scale that preserves the order between values and has well-defined intervals using real numbers is an example of a variable measured either on an interval scale, or on a ratio scale.

In practice, variables measured on interval or ratio scales are very common, if not the most common. Examples include:

* Height measured with a numerical unit of measurement (like inches or centimeters).
* Weight measured with a numerical unit of measurement (multiples and submultiples of grams, for instance).
* Time measured with a numerical unit of measurement (multiples and submultiple of seconds, for example).
* The *price* of various products measured with a numerical unit of measurement (like dollars, pounds, etc.).



**Mission 7: The Difference between Ratio and Interval Scales**

What sets apart ratio scales from interval scales is the nature of the zero point.

On a ratio scale, the zero point means no quantity. For example, the Weightvariable is measured on a ratio scale, which means that 0 grams indicate the absence of weight.

On an interval scale, however, the zero point doesn't indicate the absence of a quantity. It actually indicates the presence of a quantity.

To exemplify this case using our data set, we've used the Weight variable (measured on a ratio scale), and created a new variable that is measured on an interval scale. The new variable describes by how many kilograms the weight of a player is different than the average weight of the players in our data set. Here's a random sample that includes values from the new variable named Weight\_deviation:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name | Weight | Weight\_deviation |
| 35 | Clarissa dos Santos | 89.0 | 10.021127 |
| 3 | Alex Montgomery | 84.0 | 5.021127 |
| 111 | Renee Montgomery | 63.0 | -15.978873 |
| 85 | Layshia Clarendon | 64.0 | -14.978873 |
| 128 | Sugar Rodgers | 75.0 | -3.978873 |

If a player had a value of 0 for our Weight\_deviation variable (which is measured on an interval scale), that wouldn't mean the player has no weight. Rather, it'd mean that her weight is exactly the same as the mean. The mean of the Weight variable is roughly 78.98 kg, which means that the zero point in the Weight\_deviation variable is equivalent to 78.98 kg.

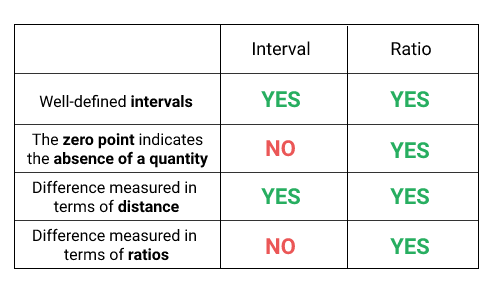
On the other side, a value of 0 for the Weight variable, which is measured on a ratio scale, indicates the absolute absence of weight.

Another important difference between the two scales is given by the way we can measure the size of the differences.

On a ratio scale, we can quantify the difference in two ways. One way is to measure a *distance* between any two points by simply subtracting one from another. The other way is to measure the difference in terms of *ratios*.

For example, by doing a simple subtraction using the data in the table above, we can tell that the difference (the distance) in weight between Clarissa dos Santos and Alex Montgomery is 5 kg. In terms of ratios, however, Clarissa dos Santos is roughly 1.06 (the result of 89 kg divided by 84 kg) times heavier than Alex Montgomery. To give a straightforward example, if player A had 90 kg and player B had 45 kg, we could say that player A is two times (90 kg divided by 45 kg) heavier than player B.

On an interval scale, however, we can measure meaningfully the difference between any two points only by finding the distance between them (by subtracting one point from another). If we look at the weight deviation variable, we can say there's a difference of 5 kg between Clarissa dos Santos and Alex Montgomery. However, if we took ratios, we'd have to say that Clarissa dos Santos is two times heavier than Alex Montgomery, which is not true.



Instructions

Examine the various variables of the data set, and find the ones that are measured on an interval or ratio scale.

* For the variables measured on a interval scale, add their names as a stringto a list named interval. Sort the list alphabetically.
* For the variables measured on a ratio scale, add their names as a string to a list named ratio. Sort the list alphabetically.
* We've also added the Weight\_deviation variable to the dataset, so make sure you include that one too in one of the lists.

If you need to consult the documentation of the data set, you can look [here](https://www.kaggle.com/jinxbe/wnba-player-stats-2017), and [here](https://www.basketball-reference.com/about/glossary.html).

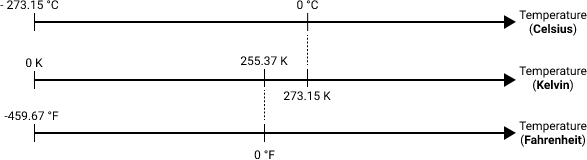
**Mission 8: Common Examples of Interval Scales**

In practice, variables measured on an interval scale are relatively rare. Below we discuss two examples that are more common.

Generally, points in time are indicated by variables measured on an interval scale. Let's say we want to indicate the point in time of the first manned mission on the Moon. If we want to use a ratio scale, our zero point must be meaningful and denote the absence of time. For this reason, we'd basically have to begin the counting at the very beginning of time.

There are many problems with this approach. One of them is that we don't know *with precision* when time began (assuming time actually has a beginning), which means we don't know how far away in time we are from that zero point.

To overcome this, we can set an arbitrary zero point, and measure the distance in time from there. Customarily, we use the [Anno domini](https://en.wikipedia.org/wiki/Anno_Domini) system where the zero point is arbitrarily set at the moment [Jesus](https://en.wikipedia.org/wiki/Jesus) was born. Using this system, we can say that the first manned mission on the Moon happened in 1969. This means that the event happened 1968 years after Jesus' birth (1968 because [there's no year 0 in the Anno domini system](https://en.wikipedia.org/wiki/Year_zero)).



If points in time are measured on an interval scale, we need to avoid quantifying the difference in terms of ratio. For instance, it's not true that twice as much time has passed from year 1000 to year 2000.

Another common example has to do with measuring temperature. In day to day life, we usually measure temperature on a Celsius or a Fahrenheit scale. These scales are examples of interval scales.

0°C or 0°F are arbitrarily set zero points and don't indicate the absence of temperature. If 0°C or 0°F were meaningful zero points, temperatures below 0°C or 0°F wouldn't be possible. But we know that we can go way below 0°C or 0°F.

If yesterday was 10°C, and today is 20°C, we can't say that today is twice as hot as yesterday. We can say, however, that today's temperature is 10°C more compared to yesterday.

Temperature can be measured on a ratio scale too, and this is done using the Kelvin scale. 0 K (0 degrees Kelvin) is not set arbitrarily, and it indicates the lack of temperature. The temperature can't possibly drop below 0 K.

**Mission 9: Discrete and Continuous Variables**

Previously in this mission we divided variables into two big categories: quantitative and qualitative. We've seen that quantitative variables can be measured on ordinal, interval, or ratio scales. In this screen, we zoom in on variables measured on interval and ratio scales.

We've learned that variables measured on interval and ratio scales can only take [real numbers](https://en.wikipedia.org/wiki/Real_number) as values. Let's consider a small random sample of our data set and focus on the Weight and PTS (total points) variables, which are both measured on a ratio scale.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name | Weight | PTS |
| 77 | Kayla Thornton | 86.0 | 32 |
| 16 | Asia Taylor | 76.0 | 31 |
| 80 | Kia Vaughn | 90.0 | 134 |
| 137 | Tierra Ruffin-Pratt | 83.0 | 225 |
| 12 | Amanda Zahui B. | 113.0 | 51 |

The first two players scored 32 and 31 points, respectively. Between 32 and 31 points there's no possible intermediate value. Provided the measurements are correct, it's impossible to find a player having scored 31.5 or 31.2 points. In basketball, players can only score 1,2 or 3 points at a time, so the points variable can only be expressed in integers when measured on an interval or ratio scale.

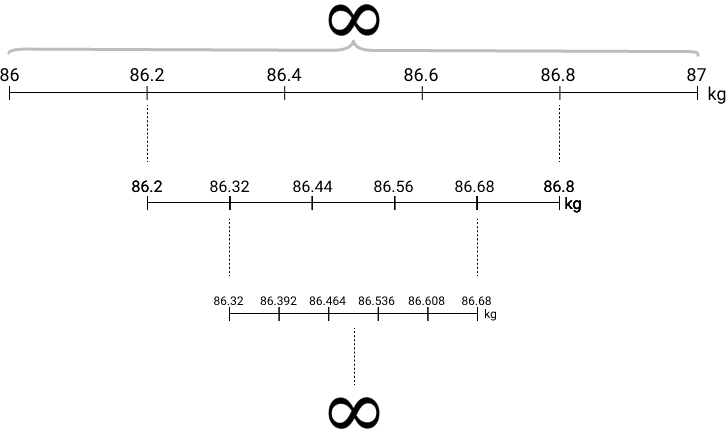
Generally, if there's no possible intermediate value between any two adjacent values of a variable, we call that variable discrete.

Common examples of discrete variables include counts of people in a class, a room, an office, a country, a house etc. For instance, if we counted the number of people living in each house of a given street, the results of our counting could only be integers. For any given house, we could count 1, 3, 7, 0 people, but we could not count 2.3 people, or 4.1378921521 people.

In the table above, we can also see that the first player weighs 86 kg, and the second 76 kg. Between 86 kg and 76 kg, there's an infinity of possible values. In fact, between any two values of the Weight variable, there's an infinity of values.

This is strongly counter-intuitive, so let's consider an example of two values that are relatively close together: 86kg and 87kg. Between these values we can have an infinity of values: 86.2 kg, 86.6 kg, 86.40 kg, 86.400001 kg, 86.400000000000001 kg, 86.400000000000000000000000000000000000000000001 kg, and so on.

In the diagram below we consider values between 86 and 87 kg, and break down the interval in five equal parts. Then we take two values (86.2 and 86.8) from the interval 86 - 87, and break down the interval between these values (86.2 and 86.8) in five equal parts. Then we repeat the process for the interval 86.2 - 86.8. In fact, we could repeat the process infinitely.



In practice, we limit ourselves to rounding the weights to a couple of decimal places either for practical purposes or because the instruments we use to measure weight are imperfect.

Generally, if there's an infinity of values between any two values of a variable, we call that variable continuous.

Whether a variable is discrete or continuous is determined by the underlying nature of the variable being considered, and *not* by the values obtained from the measurement. For instance, we can see in our data set that height only takes integer values:

print(wnba['Height'].head())

0 183

1 185

2 170

3 185

4 175

Name: Height, dtype: int64

This doesn't make the Height variable discrete. It just tells us that the height is not measured with a great degree of precision.

Instructions

For every variable, indicate whether is continuous or discrete.

* In the code editor, we've already extracted for you the names of the variables that are measured on ratio and interval scales. Every variable name is registered as a dictionary key.
* If a variable is discrete, then assign the string 'discrete' to its corresponding dictionary key.
* If a variable is continuous, then assign the string 'continuous' to its corresponding dictionary key.

If you get stuck with any variable, take any two values and try to determine whether there's an infinity of possible values between them. If you need to consult the documentation of the data set, you can look [here](https://www.kaggle.com/jinxbe/wnba-player-stats-2017), and [here](https://www.basketball-reference.com/about/glossary.html).

**Mission 10: Real Limits**

Let's consider these ten rows where players are recorded as having the same weight:

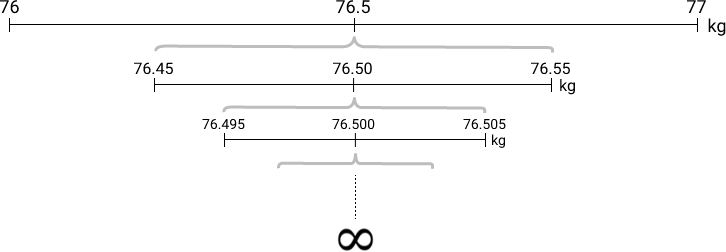
|  |  |  |
| --- | --- | --- |
|  | Name | Weight |
| 9 | Allison Hightower | 77.0 |
| 19 | Breanna Stewart | 77.0 |
| 21 | Bria Holmes | 77.0 |
| 33 | Chelsea Gray | 77.0 |
| 56 | Glory Johnson | 77.0 |
| 65 | Jessica Breland | 77.0 |
| 70 | Kaela Davis | 77.0 |
| 102 | Nia Coffey | 77.0 |
| 117 | Seimone Augustus | 77.0 |
| 132 | Tamera Young | 77.0 |

Do all these players really have the *exact* same weight? Most likely, they don't. If the values were measured with a precision of one decimal, we'd probably see that the players have different weights. One player may weigh 76.7 kg, another 77.2 kg, another 77.1 kg.

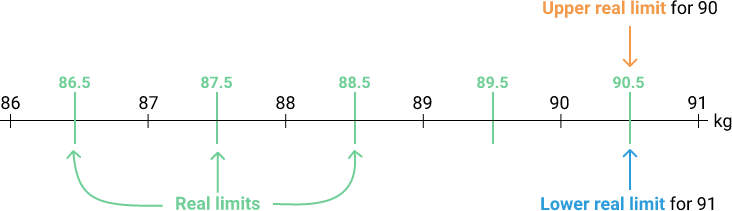
As an important parenthesis here, the weight values in the table above are all 77.0, and the trailing zero suggests a precision of one decimal point, but this is not the case. The values are automatically converted by pandas to float64because of one NaN value in the Weight column, and end up with a trailing zero, which gives the false impression of one decimal point precision. So a player was recorded to weigh 77 kg (zero decimals precision), not 77.0 kg (one decimal precision).

Returning to our discussion, if we measure the weight with zero decimals precision (which we do in our data set), a player weighing 77.4 kg will be assigned the same weight (77 kg) as a player weighing 76.6 kg. So if a player is recorded to weigh 77 kg, we can only tell that her actual weight is somewhere between 76.5 kg and 77.5 kg. The value of 77 is not really a distinct value here. Rather, it's an interval of values.

This principle applies to any possible numerical weight value. If a player is measured to weigh 76.5 kg, we can only tell that her weight is somewhere between 76.45 kg and 76.55 kg. If a player has 77.50 kg, we can only tell that her weight is somewhere between 77.495 kg and 77.505 kg. Because there can be an infinite number of decimals, we could continue this breakdown infinitely.



Generally, every value of a *continuous* variable is an interval, no matter how precise the value is. The boundaries of an interval are sometimes called real limits. The lower boundary of the interval is called lower real limit, and the upper boundary is called upper real limit.



In the figure above we can see for example that 88.5 is halfway between 88 and 89. If we got a measurement of 88.5 kg in practice, but we want only integers in our dataset (hence zero decimals precision), you might wonder whether to assign the value to 88 or 89 kg. The answer is that 88.5 kg is *exactly* halfway between 88 and 89 kg, and it doesn't *necessarily* belong to any of those two values. The assignment only depends on how you choose to round numbers: if you round up, then 88.5 kg will be assigned to 89 kg; if you round down, then the value will be assigned to 88 kg.

Instructions

Find the real limits for five values of the BMI (body mass index) variable.

* We've already extracted the first five BMI values in the data set and rounded each off to a precision of three decimal places. We stored the values as dictionary keys in a dictionary named bmi.
* For every BMI value write its real limits in a list and make the list a dictionary value that should correspond to the right dictionary key. The lower real limits should come first in each list. For example:

bmi = {20: [19.5, 20.5],

21: [20.5, 21.5],

23: [22.5, 23.5],

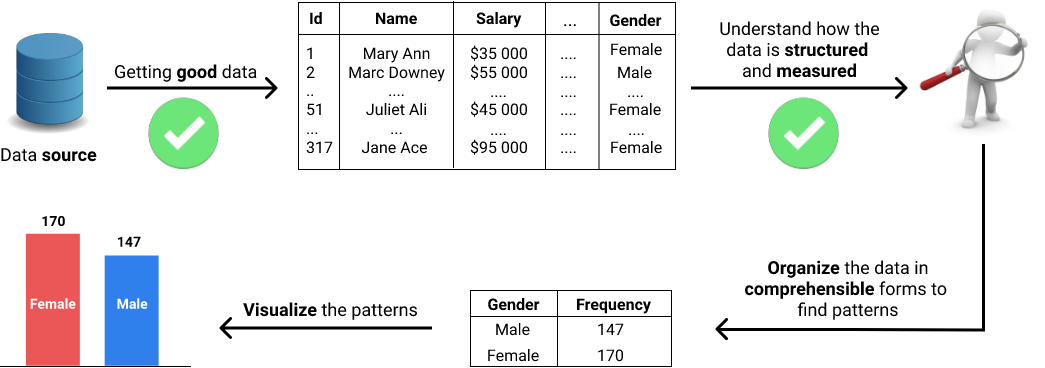
24: [23.5, 24.5],

22: [21.5, 22.5]}

**Mission 11: Next Steps**

In this mission, our focus was on understanding variables, the structural parts of a data set. We've seen that variables can be either quantitative or qualitative, and, depending on that, they can be measured on different scales.

We completed one more step in the workflow we follow throughout this course.



Next in this course, we'll learn how to organize data in comprehensible forms to find patterns.

**Mission 12: Takeaways**

Concepts

### VARIABLES

* A property whose values can vary from individual to individual is called a variable. For instance, *height* is a property whose value can vary from individual to individual — hence *height* is a variable.
* Variables can be divided into two categories:
  + Quantitative variables, which describe a *quantity*. Examples include: height, age, temperature, time, etc.
  + Qualitative variables, which describe a *quality*. Examples include: name, team, t-shirt number, color, zip code, etc.

### SCALES OF MEASUREMENT

* The system of rules that define how a variable is measured is called scale of measurement. We learned about four scales of measurement: nominal, ordinal, interval, and ratio.
* The nominal scale is specific to qualitative variables. If a variable is measured on a nominal scale:
  + We can tell whether two individuals are different or not.
  + We can't tell the direction of the difference.
  + We can't tell the size of the difference.
* The ordinal scale is specific to quantitative variables. If a variable is measured on an ordinal scale:
  + We can tell whether two individuals are different or not.
  + We can tell the direction of the difference.
  + We can't tell the size of the difference.
* Interval and ratio scales are both specific to quantitative variables. If a variable is measured on an interval or ratio scale:
  + We can tell whether two individuals are different or not.
  + We can tell the direction of the difference.
  + We can tell the size of the difference.
* For an interval scale, the zero point doesn't mean the absence of a quantity — this makes it impossible to measure the difference between individuals in terms of ratios. In contrast, for a ratio scale, the zero point means the absence of a quantity, which makes it possible to measure the difference between individuals in terms of ratios.
* Variables measured on an interval and ratio scales can be divided further into:
  + Discrete variables — there's no possible intermediate value between any two adjacent values of a discrete variable.
  + Continuous variables — there's an infinity of values between any two values of a continuous variable.