**Part 1 [0.2 points each, 2.0 points total]**

Match the following terms. You may use the items

|  |  |  |  |
| --- | --- | --- | --- |
| a) Nominal | b) Ordinal | c) Interval | d) Ratio |

\_\_a\_\_ 1) Zip code

\_\_d\_\_ 2) Number of Instagram followers

\_\_c\_\_ 3) Temperature Fahrenheit

\_\_c\_\_ 4) Temperature in Kelvin

\_\_b\_\_ 5) Level of Degree: Associate, Bachelor, Master, Doctoral

\_\_a\_\_ 6) Car Model (Ford, Nissan, Toyota, Honda, Chrysler, Jaguar, BMW)

\_\_a\_\_ 7) Months of the year

\_\_a\_\_ 8) Political Orientation (liberal, neutral, conservative)

\_\_c\_\_ 9) SAT Scores

\_\_d\_\_ 10) Weight

**Part 2**

Go onto the internet and find a dataset. It may be on any topic, but it must be in a format you can open in Excel or the R format, .Rdata. It must also have some list describing the different variables within the dataset. Download this data and dataset metadata, then open the file and explore it.

Submit a document that includes the following:

* + - 1. The name of your dataset. This can be a summarized name, in case the actual file does not make clear sense.

Average City Temperatures Across Various Regions from 1995-2020

1. Where you found the data and who created it. If the dataset author only compiled the data from other mentioned sources, and that info is given, include the acknowledged source authors.

I found this dataset on kaggle.com, which is a free, open-source website that offers access to a wide variety of datasets and other information related to data science. The dataset was published by Sudalai Rajkumar, a Kaggle user and data scientist from India. Rajkumar noted that the data was collected by the University of Dayton in Ohio.

1. A brief description of the dataset – What sort of data does it contain (e.g. Past 5 years of housing prices in Seattle, WA)

The dataset contains the average daily temperatures from 1995-2020 of various cities across the following regions: Africa, Asia, Australia/South Pacific, Europe, and the Middle East.

1. Is the data observational or administrative? Primary or secondary?

This data is observational because it was collected by watching and recording the average daily temperatures for different cities. The data is secondary because it was originally collected and published by the University of Dayton, which was then accessed by Rajkumar and published on Kaggle. And now, I have access to the data, too.

1. Identify 1 independent variable (IV) and 1 dependent variable (DV) and post about what sort of research question you could ask using those 2 variables.

An independent variable in this dataset is the region (e.g. Africa, Asia, Europe, etc.) and a dependent variable is average daily temperatures of a city.

A research question I could ask with these variables is whether the average daily temperatures of a city are affected by the location of the region it’s in. For example, I could investigate whether the average daily temperatures in Cairo, Egypt, are caused by its location in the continent of Africa and how average daily temperatures differ for cities in other regions.

1. A link to where you found the data

<https://www.kaggle.com/sudalairajkumar/daily-temperature-of-major-cities>

When you upload this homework to ELMS, also upload the dataset and the variable list or other related documents.

**Part 3**

3a) Researchers recently measured the number of microbiota of the genus Bacteroides that a group of 60 infants had when they were one-year-old and found that the more Bacteroides microbiota, the better the infants' gross motor skills, perceptual abilities, and language development by age 2.

1. What is the independent variable?

The independent variable is the number of Bacteroides microbiota in a one-year-old infant.

1. What are the dependent variables?

The dependent variables are the levels of gross motor skills, perpetual abilities, and language development by age 2.

1. What kind of study design is this?

This is an observational study design because the researchers are not implanting the Bacteroides microbiota in the infants, it’s naturally occurring in the infants’ bodies. Instead, the researchers are observing the number of Bacteroides microbiota present and whether higher/lower counts influence an infant’s levels of gross motor skills, perpetual abilities, and language development by age 2.

1. What are two weaknesses of the study? There can be several answers to this including:

One weakness of this study is that it doesn’t seem to be accounting for any outside influences that may affect an infant’s levels of gross motor skills, perpetual abilities, and language development like activities a parent might do with their child (e.g. Use of baby walkers/bouncers and other toys, speaking to child without baby talk, etc.). Another weakness of this study is the small sample size. To get a better understanding of how the number of Bacteroides microbiota present in an infant affects development, the researchers should use a larger sample size that is also representative of different races, ethnicities, religions, family types, etc.

3b) An organization wants to know if drinking caffeinated coffee causes hyperactivity in college students. To test their research question, they offered students free coffee to students throughout the term in exchange for participating in the study. Researchers intentionally balanced participating students into groups by the student’s year, major, and gender. One group received regular coffee and the other received decaffeinated coffee. Students obtained their free coffee at will from the research stations and through the day used an app to track how hyperactive they felt.

1. What are the independent variables?

The independent variables include the free coffee provided by the researchers, the student’s year, major, and gender, and whether the coffee was regular or decaffeinated.

1. What is the dependent variable?

The dependent variable is the level of hyperactivity the research participants felt throughout the day.

1. What kind of study design is this?

This is a quasi-experimental study design because while the researchers attempted to control certain aspects of the study like whether the coffee was regular/decaffeinated and the groups students were placed into by year, major, and gender, they did not control if a student would pick up free coffee.

1. What kind of sampling method is being used?

The sampling method used is stratified random because the initial group of students used in this study were split up by the researchers based off of year, major, and gender.