DA 6823

Kilger

Module 1: Part #1 (50 points)

**The Power of Statistics + the Levels of Measurement + the Different Classes of Variables and Determining Appropriate Statistical Technique + Basic Descriptive Measures**

**General Instructions:** In your own words, answer each of the following questions - don’t copy (e.g. cut and paste) some definition out of a book word for word. This is not a group project – you are expected to complete this module on your own. You may refer to text books, online or other sources but not your fellow classmates. If you don’t understand the question, feel free to ask the instructor in class, in office hours or in an email.

1. Provide a short definition for dependent variable. (3 points)

In my mind, I think of cartesian coordinates and X’s and Y’s. My function takes X as an input, i.e., independent variable, does some magic and spits out the Y variable, i.e., the dependent variable. This the output depends on the input.

1. Provide a short definition for independent variable. (3 points)

Given the explanation in 1, the independent variable is the input to the function, or what we feed into the machine, which munches on it and spits out the answer or dependent variable.

1. Provide a short definition for control variable. (3 points)

A variable that is kept constant. Again, referring back to questions 1 and 2 and thinking about the slope intercept equation (y=mx+b), b (the y-intercept) would be the control variable, x is the independent variable and y is the dependent variable since it’s value depends on x.

1. Be able to describe the simple criteria for each of the four levels of measurement:
   1. Nominal (2 points)

Nominal, or named variables, are non-numerical variables and are used for classification. An example of nominal variables would be categories, such as male and female for gender.

* 1. Ordinal (2 points)

As the name implies, ordinal variables are ordered or ranked, e.g., first, second, third, etc.

* 1. Interval (2 points)

The level of measurement is the first, in this hierarchy, to specify a numerical value for a measurement. This value is the interval or distance between 2 other values and is held constant over the scale, i.e., the interval between 1 and 2 is the same as the interval between 6 and 7.

* 1. Ratio (2 points)

Is the fourth level in the measurement hierarchy. A ratio allows us to specify not only specify equal intervals between values, but the set of values also has a 0 reference point.

1. Provide an example of a variable for each of the four measurement levels below.
   1. Nominal (2 points)

Male/Female

* 1. Ordinal (2 points)

First, second, third, ….

* 1. Interval (be careful – be sure it is interval and not ratio!) (2 points)

Distance from 30 to 40, i.e., 10.

* 1. Ratio (2 points)
     + 5 feet
     + 50 kilograms

1. Name at least two criteria from the IDRE chart that are used in determining which statistical technique can be used in a situation. (3 points)
   1. The nature of the dependent variable, i.e., ordinal or interval
   2. The number of dependent variables
2. Briefly explain the difference between descriptive and inferential statistics. (4 points)
   * + As the name implies descriptive statistics employs means for describing the data, which may include using charts, graphs, and tables.
     + Inferential statistics looks to provide conclusions, i.e., make inferences, beyond the data itself, giving us the “bigger picture”.
3. Almost every statistical technique you will come across has some sort of assumptions – even non-parametric statistics.
   1. Name one benefit of that assumptions of a test provides you (2 points)

By making assumptions, we can determine the types of tests that make sense for a particular dataset.

* 1. Name one cost that assumptions of a test carry (2 points)

If our assumptions about the dataset are incorrect, then our results will most likely be incorrect as well.

1. What happens if you violate the assumptions of a statistical test? Do the statistical police come and arrest you? (4 points)

Violations of assumptions impacts our ability to trust the test results, and, no, the statistics police will not come, but our colleagues may not ask for our help if they cannot trust our statistics.

1. Using the IDRE chart, suggest the appropriate statistical test for each of the following business cases
   1. As a maker of colored contact lenses, you think that there may be relationship between the color of the contact lenses purchased and the gender of the purchaser. (2 points)
   2. As an auctioneer of fine art, you think that there may be a different between the price paid for a piece of art between men and women. (2 points)
   3. You want to better understand how different versions and price mixes of your product – the Vegematic – have on the number of product sold. You hypothesize that color of product, price, region of the country (North, South, East, West), gender of purchaser, household income of purchaser have an effect on the number of pieces sold. You may also want to make some predictions about how many products would be sold under various levels of these variables. (2 points)
   4. As publisher of the popular magazine Rabbit Times, you think that there may be a relationship between the number of pages in the magazine and the number of copies of that issue sold. How do you find out the direction and how strong this relationship might be? (2 points)
   5. You are the maker of FelineHair – a hair growing drug for hairless cats. You want to test your drug against three other drugs to see which one grows the most hairs on the cats in the experiment. You also want to see if there are other differences in the effectiveness depending upon the gender of the cat and what color coat the cat has. You end up with a drug (4) x cat gender (2) by cat coat color (black, white, brown) experimental design. What analysis technique would you use for this experiment? (2 points)