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**

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

The dependent variable also called response variable is the variable we want to test, study or measure and whose value depends on one or more other variables/factors.

Its value changes as a result of change in those variables/factors.

For example, test score in an exam could be a dependent variable whose value may depend on several factors being, time spent on preparation, health condition, amount of time spent sleeping the night before etc.

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

Any variable whose value affects the value of other variable is called independent variable. Also known as explanatory/predictor variable, it can be modified and manipulated to measure the value of a dependent variable.

For example, the time spent on preparation could affect the test score hence could be considered as independent variable.

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

A control variable is held constant in order to assess the relationship between dependent and independent variables. If it is not held constant, we wouldn’t know if the change in dependent variable is due to independent or control variable

For example, to see how the amount of light received effects the plant growth, factors like water and fertilizer levels which might influence the plant growth need to be held constant.

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

Nominal variable also known as categorical variable has two or more categories that do not have any ordering among them. When numbers are used as tags to classify the variables, calculations can’t be performed on these numbers as they have no quantitative significance.

* 1. Ordinal (2 points)

Ordinal variable is same as nominal variable that has two or more categories but also can be ranked/ordered. The distance/interval between categories doesn’t have any meaning. Unlike interval variables, the distance between two levels cannot be assumed the same as the distance between two other levels.

* 1. Interval (2 points)

The order as well as distance between values for interval variable is meaningful.

Its value can go below 0. Categories are equidistant from each other. Ratios don’t make sense in interval variable

* 1. Ratio (2 points)

Ratio variable has absolute zero point that is meaningful. When the value of a ratio variable is 0 then there is none of that variable. It can be put into categories like nominal variable, be ordered like ordinal variable, categories are equidistant from each other as in interval variable and also the ratio of the levels makes sense.

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

Marital status of a person can be one of different possible categories:

* Single
* Married
* Divorced
* Widowed
  1. Ordinal (2 points)

Level of Educational experience of a person:

* 1. Elementary school graduate
  2. High School graduate
  3. Bachelor’s Degree
  4. Master’s Degree
  5. Interval (be careful – be sure it is interval and not ratio!) (2 points)

Celsius scale of temperature: Has meaningful value below 0 and distance between 10 and 20 degrees Celsius is same as distance between 20 to 30.

30 is always higher than 20 which makes it ranked.

* 1. Ratio (2 points)

Weight is a ratio variable. 0 weight means no weight at all. Someone weighing 80Kgs is twice as heavy as someone weighing 40Kgs.

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)

* Nature of Independent variables
* Number of dependent variables
* Nature of dependent variables (nominal ordinal or interval)
* Independence or dependence of groups or observations

1. Briefly explain the difference between descriptive and inferential statistics. (4 points)

* Descriptive statistics as the name suggests, describes or summarizes the sample that is available to us, in a meaningful way.

Inferential statistics on the other hand helps us draw conclusions/make generalizations about the entire population based on the sample drawn from the population. It is important that sample accurately represents the entire population. We achieve this through random sampling

* Descriptive Statistics analyses the data that is already known

Inferential Statistics tries to extrapolate about the population that extends beyond the data available

* Descriptive Statistics uses statistical measures like central tendency (mean, mode), dispersion (range or standard deviation), skewness

Inferential Statistics uses methodologies such as hypothesis tests, confidence intervals and regression analysis

* Descriptive Statistics uses graphical representation of data through graphs, charts and tables

Inferential statistics uses probability (p-value) to make conclusions

1. 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)

The assumptions help to classify the data properly or draw conclusions accurately and efficiently in order to avoid false positives or false negatives which is critical in medicine/health industry, etc.

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

We need to conduct various tests to check for assumptions and make transformations if some assumption does not hold true. Mostly is real time scenarios all the assumptions may not hold true and it takes time to analyze and come to a conclusion

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

When assumptions are violated the results could be misleading or completely erroneous.

Critically would depend on the purpose of research/study. Violating assumptions while conducting cancer research for example could result in incorrect false positives or false negatives which could prove fatal

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)

Dependent Variable – Lens Color, Independent variable - Gender

Chi-square test /Fisher’s exact test

* 1. 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)

Dependent variable - Price paid, Independent Variable-Gender

If the residuals distribution of price paid is normal – 2 independent sample t-test

If the residuals distribution is not normal – Wilcoxon-Mann Whitney test

* 1. 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)

Independent variable – color of product, price, region of the country (North, South, East, West), gender of purchaser, household income of purchaser

Dependent variable - Number of products sold (interval variable)

Appropriate Statistical test - ANCOVA/multiple regression

* 1. 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)

Dependent variable – Number of copies cold, Independent – Number of pages

Statistical test – Simple linear regression. The sign of coefficient estimate gives the direction. Correlation coefficient from correlation matrix/plot tells how strong the relationship is.

* 1. 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)

Independent variable – coat color, Gender, drug type

Dependent variable – hair volume

Statistical test - Factorial ANOVA