Statistics essay sample

Environment, Water

Every event in our life depends on many factors. Some of these factors are predictable but the other can affect randomly. This makes our world stochastic and full of random values. Therefore, every measurement we take, every experimental data gives us not the precise but the most probable result. To estimate obtained results we use statistics methods. It gives us possibility to plan studies and experiments; to obtain, to summarize and to analyze results as well as to draw conclusions or to make predictions. On the first step we need to choose appropriate data for analysis. Sometimes researcher has a lot of information that is nearly impossible to analyze fast. For instance, it is impossible to interview fast million people about their opinion of some political or environmental issue (Yue 81). In this case it is possible to take small part of data that will represent characteristics of all population. Such a group is called sample and the process is called sampling.

There are two main groups of sampling methods. These groups are called probability (or random) sampling and non-probability (non-random) sampling. Using probability sampling we imply that every subject has a chance to be selected from the population. In non-probability sampling researcher utilizes other criteria for choosing. It can be possibility to get information, specialization on separate group of data or similar characteristics of the objects (Mann 55).

Four types of methods are used in random sampling. These types include simple random sampling, systematic random sampling, stratified random sampling, and cluster sampling. When researcher uses simple sampling, there is equal probability for each object to be chosen. In systematic

sampling, we use random starting point and select members at equal intervals. For instance we can take every tenth probe of water starting from probe number 5 (5, 15, 25). During stratified sampling, researcher assigns data to separate groups and then chooses objects from each group (Mann 58). The number of chosen objects is proportional to the number of objects in group. For example, for the territory that consists of 70 % urban land and 30 % agricultural land 70 % of probes should be taken from urban land and 30 % of probes should be taken from agricultural land. Cluster sampling foresees assigning each object to separate group and random selection of cluster to sample group. Random sampling methods permit to eliminate effect of researcher opinion during selection process (Mann 60).

However, we may use opinion for planning separate types of estimation. Non-probability methods are used for this purpose. Our sampling procedure can be bases on availability of information for the study (convenience sampling) or on our judgment to select object with similar characteristics (purposive sampling). We can also select data basing on additional information from certain person(s) (snowball) or to define subgroups and to select object from each (quota). Such sampling methods give possibility for the researcher to generalize studied properties for specified group (Mann 62).

After sampling, we can analyze data and generalize obtained results for population or subgroup. The main purpose of our analysis is to precise (or most probable) value of certain characteristics. We can define mass, concentration, area, content of pollutant in water or soil samples as well as data obtained from interview. All data have the different value. The most

probable value of the sample will be mean value. We calculate it dividing sum of all variables on the number of variables. For example, we can calculate average content of substance in water (Yue 217). Afterward, it is possible to calculate range, variance and standard deviation. Range is calculated as difference between maximum and minimum value in the sample. To calculate variance and standard deviation the following formula are used (Mann 92)

$s2= (x-x)2n-1$

$s= (x-x)2n-1$

where x is value of variable, n is number of variable and x is mean value. The latter characteristics show how random value differs from mean value.

It is also possible to build graph number of variables with certain value vs. this value. This graph shows distribution of variables near mean value. Most of the random variables are normally distributed. Area under the obtained curve of normal distribution gives the probability that variable will have certain value (Mann 306).

After obtaining measures of variation, we can test hypothesizes. For example, we can test hypothesizes that content of substance in the water samples is greater, less, equal or not equal to the certain value. To test hypothesis on the first stage we formulate null (true) and alternative (false) hypothesizes. The next step is to define the level of significance. We can set that with probability of 10 % content of substance in the water is greater than 10mg/l. When the probability to obtain such result is greater than 10 %, our hypothesis is true.

The third step of testing is calculation the test statistics. We may use Z-score

for this purpose. Z-score is calculated using the following formula

$z = x - xs$

where s is standard deviation (Mann 390).

The final step is finding corresponding to the z-score probability and comparing. We can reject or retain null hypothesis. In other word, we can confirm that hypothesis is true or false.

Procedures of sampling, measuring of variation and hypothesis testing can be used in many fields of study. We can use it, for instance, in chemistry or environmental studies. We may define representative group of probes that will show the content of pollutant in the water object. After sampling, it is possible to calculate mean value, variance and standard deviation. The mean value is the most probable result of investigation. Afterward, we can test hypothesizes concerning whole data population.

Statistics gives us possibility to analyze data and to get more precise vision of the processes or facts. Using statistics methods, we can avoid errors and to eliminate the influence of unpredictable factors. It also gives us possibility to save time while analyzing the large amounts of information.

Works cited

Mann, Prem S. " Introductory Statistics." John Wiley & Sons (2010): 1-438. Print

Rong, Yue. " Practical Environmental Statistics and Data Analysis." ILM Publications (2011): 1-113. Print.