Interpretation of Statistics in Wildlife Research

ABSTRACT

There is a huge amount of research happening on wildlife population but they add little value in real life. Most of the studies does not give correct conclusions because of many reasons. One must be choosing inappropriate sample size. Failed to apply study population inferences on target population because of various factors. Most of the times it is lack of interpretation of data. Lack of understanding of statistical and biological significance on information and its importance in hypothesis testing. Not meeting assumptions before a test is conducted. Misjudging extreme data points in the data present.

This paper describes how statistical hypotheses are often viewed and then explains how it has to be viewed. Populations from which inferences are drawn should be clearly defined and conclusions should be limited to the range of data analyzed. Model Selection, Randomization are discussed. Bayesian approach to hypothesis testing is also explain in the paper.

This paper describes how statistical hypotheses are often viewed and then explains how it has to be viewed. There is a huge amount of research happening on wildlife population but they add little value in real life. Most of the studies does not give correct conclusions because of many reasons. Lack of understanding of statistical and biological significance on information and its importance in hypothesis testing. Model Selection, Randomization are discussed. Bayesian approach to hypothesis testing is also explained in the paper.

Introduction

There is a huge amount of research happening in the The Journal of Wildlife  
Management from nearly 0 to 87% between1938 and 1978 (Thomas C. Tacha W. D., 1982).As per Kish (1969:7) definition of a population should include content, units, time frame, and extent. Only 29% of statistical study was able to get information which included all these 4 criteria. (Thomas C. Tacha W. D., 2006)

In any research the population is considered in 2 levels. One is study population which gives statistical inferences and the second is target population which is used to apply inferences which obtained from study population. Of the articles present in The Journal of Wildlife Management, more than 42% of them were able to proper inferences from study population but failed to apply those conclusions properly to target population. A major ornithological journal reported a study, they observed that only 3 birds were marked in 150 sqkm study area for 3 years. As per this data they gave conclusions for the entire breeding biology of the species. The study population is 150 sqkm for 3 years, but the target considered is the entire population present in all states. (Thomas C. Tacha W. D., 2006)

Central Tendency Measurements

There are several measures of tendency available for population mean, median and mode. Based on the dataset and type of prioritizing we can choose the measure. If Parametric testing such as hypothesis testing has to done, then we use mean. For the non-parametric tests median, mode, range and interquartile range are to be considered.

Biological vs Statistical Significance

Many times, authors appear to be confused with statistical and biological significance. Before coming to conclusion of results authors should understand the relation between biological significance and statistical significance. Let us consider an example for better understanding.

Example: In a wildlife research there is a biological significance of 5% but not for statistical significance. When 2 forest areas have 67 and 59 deer each. Ten sample counts of each area provide 69(10) and 49(10), gets observed significance level 0.06 .Most of the times authors conclude that there is on difference in deer and conclude .There is estimated difference of 20 significant at p=0.06 level. Sometimes the conclusions can be biologically logical but statistically there are not significant.

The biological significance and statistical significance have high correlation when sample sizes are large, and power is high. However, the biologically trivial differences may not be statistically significant when power is low. The difference in biological and statistical significance is because of sample size or insufficient design. (Thomas C. Tacha W. D., 2006)

Extreme data and its effect on the data spread

Most of the cases it is very difficult to address all the assumptions associated with statistical models. It is very important to address the assumptions which can violate the conclusion. If the assumptions which can affect the results are violated, then we should use alternative methods. There are cases where assumptions cannot be tested, and the violations are suspected, in that case it is to be reported in the research paper or article. Most of the statistical data of wildlife population present is considered as normal distribution (t test). This distribution can be affected when there is an extreme point in the data, especially for t tests where the sample size is small and one extreme point in data can change the spread and distribution. So, it is a good practice to plot the data and look for normal distribution before starting any parametric or non-parametric tests. Even though parametric tests have more power in detecting difference, it progressively misleads as the data moves away from the normality. (Thomas C. Tacha W. D., 2006)

Let us consider a study where sample consists of 10 males and 3 females. What samples are extreme data when null hypothesis is balanced sex ratio ? The outcomes with extreme data are 11 males and 2 females, 12 males and 1 female, 13 males and 0 females for the sample size 13.

The researcher has decided to stop sampling when he gets 10 males. In this case the outcomes of the extreme data are 10 males and 2 females, 10 males and 1 female, 12 males and 0 female. Similarly the data collected when sampling stops for 3 females . The outcomes are 11 males and 3 females, 12 males and 3 females, 13 males and 3 females which are extreme. Thus a researcher can get proper sampling when he decides to stop sampling until the difference between males and females was 7. (Johnson, The Insignifance of Statistical Significane Testing, 1999)

Applying Hypotheses testing on Wildlife situations

In wildlife research hypotheses test is much useful to know the sign of the effect. The test cannot give the magnitude of the effect. Let us consider a wildlife study , the hypotheses test is to check if the survival rate of mallard increases when sport hunting is stopped in North America. Most of the researchers believed that the hypotheses test will be true but to what extent. Does the survival rate increase worthwile when compared to the loss in recreational opputunities. In order to come to conclusion other than knowing the increase or decrease of survival we should also know how big the increase will be. Because stopping recreational activity is a big and costly affair which can have social implications as well.

Randomization

In wildlife research it is very much important to have randomization in sample data . Let us take an example of vegetation in a forest to explain in detail. If a researcher wants to know characteristics of vegetation in 10 hectars field. You will take 8 quadrants and measure vegetation in that area. The conclusions we draw from that will project the entire area. There is a chance that the 8 quadrants selected can be with in the same area and they might be very different from the entire field. In these cases it is always better to take random samples multiple times which can increase our chance of efficiency with the hypotheses test.

Model Selection

Based on the data present Bruhman and Anderson provided overview of several models. In any research one should not select the best model and then proceed with it. With the information we have, different strengths of evidence can be provided for each model of the same data. Rather on deciding or making conclusion on one model which is strongly supported by the data. It is always advisable to consider a set models and weight the strength of evidences. (Johnson, The Insignifance of Statistical Significane Testing, 1999)

Bayesian Approaches

Just as parametric modelling, Bayesian modelling is a statistical model with an extra step. The main trouble researchers face with statistical analyses is that some target population is not well understood. Also, some assumptions of it are ignored. This uncertainty can be modelled with Bayesian modelling in a Probabilistic way.

For the information present we have N possible mutually exclusive hypotheses that cover all scenarios. The Bayes theorem is based on the fact that the conditional probability P (Hi | E) i.e., the event E occurs under hypothesis Hi. (M. Stein, 2013)

https://latex.codecogs.com/gif.latex?P%28H_%7Bi%7D%7CE%29%3D%20P%28H_%7Bi%7D%5Ccap%20E%29%20/P%28E%29

Since the hypotheses *Hi* are mutually exclusive and cover all possible situations

https://latex.codecogs.com/gif.latex?P%28H_%7Bi%7D%7CE%29%3D%20P%28H_%7Bi%7D%5Ccap%20E%29%20/%20%5Csum_%7Bi%3D1%7D%5E%7BN%7DP%28H_%7Bi%7D%5Ccap%20E%29

Let prior probabilities of each of these hypotheses *P*0(*Hi*)

https://latex.codecogs.com/gif.latex?P%28E%20%5C%26%20H_%7Bj%7D%29%20%3D%20P%28E%20%7C%20H_%7Bj%7D%29%20%B7%20P_%7B0%7D%28Hj%29

Thus,

https://latex.codecogs.com/gif.latex?P%28H_%7Bi%7D%7CE%29%20%3D%20P%28E%20%7C%20H_%7Bi%7D%29%20%B7%20P_%7B0%7D%28H_%7Bi%7D%29%20/%5Csum_%7Bj%3D1%7D%5E%7BN%7D%20P%28E%7CH_%7Bj%7D%29%20P_%7B0%7D%28H_%7Bj%7D%29

When the uncertainty is given parameter θ ,then Bayesian treats θ as the random variable of distribution.

https://latex.codecogs.com/gif.latex?g%28%5Ctheta%20%7Cx%29%3Df%28x%20%7C%5Ctheta%20%29.g%28%5Ctheta%29%5Cint%20f%28x%7C%5Cnu%20%29.g%28%5Cnu%20%29%20d%5Cnu

Power Analysis

Whenever research is done on wildlife population it is very important to plan power analysis to determine the sample size. We can calculate power when sample size, significance level are given. One of the goals of any research in wildlife population is to increase the efficiency and minimize the experimental error. In any technique as the error reduces then the statistical power increases.

SUMMARY

In this paper we give suggestions on what biologists or researchers should know before doing hypothesis testing. It is acknowledged that all statistical null hypotheses are rejected even before researchers gather the data or any tests are performed. (Johnson, 2002, pp. 272-276) Researchers should not rely on completely on a single study conducted in one year ,instead they should consider results from multiple researchers using different methods.

Biological hypotheses address fundamental and global details whereas statistical hypotheses address is more local and concentrate on single population system. Also, researchers should think about biological hypotheses more often than statistical hypotheses

# Bibliography

Johnson, D. H. (1999). The Insignifance of Statistical Significane Testing. In *The Journal of Wildlife Management* (pp. 763-772).

Johnson, D. H. (2002). The Role of Hypotheses testing on Wildlife Science. In D. H. Johnson, *The Journal of WildLife Management* (pp. 272-276).

M. Stein, a. M. (2013). Bayesian Approach for Inconsistent Information. *Bayesian Approach for Inconsistent Information*, 96-111.

Thomas C. Tacha, W. D. (1982). Use and Interpretation of Statistics in Wildlife Journals.

Thomas C. Tacha, W. D. (2006). USE AND INTERPRETATION OF STATISTICS IN WILDLIFE JOURNALS. *USE AND INTERPRETATION OF STATISTICS IN WILDLIFE JOURNALS*, 355-362.