Popular Article

Evaluating the Randomness of Observations at the Veterinary Clinical Complex, Jabalpur: A Run Test Application

Diwakar Mishra[1], Ankur Tirkey[1], Mayank Meena[1], Muskan Yadav[1], Mohan Singh Thakur[2], Asit Jain[2], Apra shahi[2]

[1]Department of Veterinary Surgery and Radiology, College of Veterinary Science and A.H., Jabalpur - 482001, M.P., India

[2]Department of Animal Genetics and Breeding, College of Veterinary Science and A.H., Jabalpur Jabalpur - 482001, M.P., India

Corresponding author: *Diwakar Mishra, Email: diwakarmishra332@gmail.com

Abstract

Determining the randomness of clinical study data often depends on whether samples are collected using probability or non-probability methods. Probability sampling aims for truly random samples, while non-probability sampling might still produce random data but without guarantees. This article explores the use of the run test to evaluate data randomness from an ordered population, particularly in clinical settings where sequential sampling is common. The run test, a non-parametric statistical method, is used to determine whether a sequence of observations adheres to a random pattern. In this study, we applied the run test to a sequence of 20 patient observations from the Veterinary Clinical Complex (VCC) in Jabalpur, categorized as Non-Descripts (ND) and Descripts (D). Our goal was to determine if the sequence of breeds attending the VCC was randomly arranged. We developed and refined hypotheses to test for randomness: the null hypothesis (H0) assumed randomness, while the alternative hypothesis (H1) suggested a non-random arrangement. There were nine runs in the sequence when we tallied the number of runs. Using critical values for small samples from Swed and Eisenhart's table, we applied the decision rule and accepted the null hypothesis. The results confirmed that the sequence of breeds was random, showcasing the run test's effectiveness in assessing sampling randomness.

Introduction

The terms 'random' and 'non-random' are often used to describe whether data in a clinical study has been collected through a probability or non-probability sampling method. Probability sampling presumes that the sample is chosen randomly to gather data. However, even in non-probability sampling, it is still possible to collect a random set of data, though there is no guarantee that the data will be truly random. This article focuses on sample selection from an ordered population in clinical surveys, where subjects are typically selected in a 'first come, first served' manner. In such ordered populations, like patients waiting in line to see a doctor or cars lining up at a drive- thru, the sequential sampling method is commonly used. Sequential sampling, a form of non-probability sampling, is straightforward and widely applied in these contexts. However, it is crucial to examine whether the data collected through this method is truly

random, as the validity of inferences drawn using inferential statistics depends on the absence of bias in the sample data. If the sample is not randomly selected, the results may be skewed, compromising the accuracy and reliability of the study's conclusions. Therefore, careful consideration of the sampling method is essential to ensure unbiased data collection in clinical research.

Understanding the Principles and Applications of the Run Test

A run is a distinct sequence within a dataset where occurrences of one type of event are followed or preceded by occurrences of a different type or by no events. An excess or deficiency of runs may suggest that the process is not random. The run test is a statistical method used to determine whether a selection during the sampling process was made randomly from an ordered group. Since the run test is non-parametric, it may be used in a variety of scenarios without requiring the assumption of a normal distribution. For a one-sample run test, the variable should be dichotomous, while in a two-sample test, the samples need to be independent from each other. In this article, we aim to explain the methods for conducting a run test and demonstrates methods for conducting a run test and demonstrates its application through practical examples.

When evaluating the randomness of selections, a run test is very helpful as it may be used as a sign of possible bias in the selection procedure. By identifying irregularities in the sequence of events, the run test can help ensure the integrity and fairness of the sampling method, ultimately leading to more reliable and unbiased results in research studies.

Hypotheses

When using the run test, researchers must carefully select the appropriate hypothesis based on the specific aspect of randomness they wish to investigate. The run test allows for different hypotheses, each tailored to the particular research question at hand. The primary hypotheses that can be investigated are listed below:

1. Two-sided hypothesis:

a. Null hypothesis (H0): The patterns seen in the two types of observations are the result of a random process.

b. Alternative hypothesis (H1): The patterns seen in the two types of observations are not the result of a random process.

2. One-sided hypothesis (focusing on too few runs):

a. Null hypothesis (H0): The patterns observed for the two types of observations are determined by a random process.

b. Alternative hypothesis (H1): The patterns observed for the two types of observations are not determined by a random process because there are too few runs.

3. One-sided hypothesis (focusing on too many runs):

a. Null hypothesis (H0): The patterns observed for the two types of observations are determined by a random process.

b. Alternative hypothesis (H1): The patterns observed for the two types of observations are not determined by a random process because there are too many runs.

Selecting the correct hypothesis is crucial, as it guides the interpretation of the run test results. Whether the concern is about the number of runs being too high,

D	D	ND	ND	D	ND	D	D	D	D	ND	D	D	ND	ND	ND	ND	ND	D	D
1		2		3	4	5				6	7		8					9	

Table1: Calculation of run in a sequence of 20 patients attending VCC, Jabalpur

too low, or simply different from what would be expected by chance, The choiceof hypothesis affects both the analysis and the outcomes of the research.

Suppose a researcher observes a sequence of 20 patients attending Veterinary Clinical Complex (VCC), Jabalpur. The researcher aims to assess whether this sequence exhibits a random pattern depending on the breed of the subjects (ND = Non-Descripts and D = Descripts).

D DND NDDNDD D D DNDD DND NDNDNDNDD D

A two-sided test was first used by the researcher to construct the hypothesis. This hypothesis can now be revised to better align with the specific context of the study. The refined hypothesis will more accurately address the particular scenario for hypothesis testing, as detailed below:

H0: The sequence of patients' breeds attending the VCC follows a random pattern. H1: The sequence of patients' breeds attending the VCC does not follow a random pattern and is arranged non-randomly.

Next is to count the number of observations categorized as ND (Non-Descripts) and D (Descripts), and also tally the number of runs in this series.

A run is defined as a sequence of consecutive observations with the same value. In this example, the researcher will count the number of consecutive occurrences for both ND (Non-Descripts) and D (Descripts). The process for calculating the runs are outlined in table 1, which shows that the number of D (n1) is 11, ND (n2) is 9 and runs (r) is 9, respectively.

Test Measurements

The next step involves calculating the test statistic. A test statistic can be defined as a computed value obtained from sample data that gauges how closely the sample data matches the null hypothesis. This value is essential for evaluating whether the observed data supports the null hypothesis or suggests its rejection.

To determine whether to reject the null hypothesis, the test statistic's estimated value is compared to critical values derived from the expected distribution under the null hypothesis. The method for calculating test statistics and critical values varies depending on the sample size. For small samples, the number of runs is directly used to determine the test statistic, while for larger samples, an approximation method is employed.

For small samples, typically when the sample size is 20 or fewer, critical values are derived from randomness run tests. Conversely, for larger samples, critical values are calculated using an approximation formula, which generates a table of important factors.

In practice, if the sample size (n1 and n2) is 20 or less, as in this example, the test statistic is computed based on the number of runs observed. For instance, if the data

from two events show 9 runs in total, then the test statistic r is 9.

In summary, for small samples with 20 observations or fewer, the number of runs directly represents the test statistic. For larger samples, an approximation technique is applied to determine critical values. In this case, the small sample test involves analyzing 20 observations, where the number of runs is used as the test statistic, with 9 runs being the result for the given example.

Critical Value for Small Samples

The critical value can be retrieved from the table provided by Swed and Eisenhart. From that standard table the upper and lower critical value obtained for n1 = 11and n2 = 9, are 6 and 16, respectively.

Decision Rule

For a small sample, such as in this case, the decision rule based on different types of hypothesis statements can be summarized as follows:

1. Two-sided:Reject H0 if $r \leq Lc$ or $r \geq Uc$

2. One-sided (H1 indicates too few runs): Reject H0 if $r \leq Lc$

3. One-sided (H1 indicates too many runs): Reject H0 if $r \geq Uc$

In this example, since $6 \leq r = 9 \leq 16$, the decision is to accept H0. Therefore, we can conclude that there is enough evidence to claim that the sequence of breeds attending VCC had been selected randomly.

Conclusion

The run test is a valuable tool for assessing sampling randomness and upholding the integrity of data collection. By accurately applying the run test, researchers can detect potential biases in the data, which enhances the reliability and validity of clinical research findings. This ensures that the results reflect true patterns rather than random anomalies.

References

1. Aslam, M. (2023). The run test for two samples in the presence of uncertainty. Journal of Big Data, 10(1), 166.
2. Mohanty, S., Achary, A. A., & Sahu, L. (2022). Improving Suspicious URL Detection through Ensemble Machine Learning Techniques. In Society 5.0and the Future of Emerging Computational Technologies (pp. 229- 248). CRC Press.
3. Bujang, M. A., & Sapri, F. E. (2018).An application of the run test to test for randomness of observations obtained from a clinical survey in an ordered population. The Malaysian Journal of Medical Sciences: MJMS, 25(4), 146.
4. Lim, A. J., Khoo, M .B., Teoh, W. L., & Haq, A. (2017). Run sum chart for monitoring multivariate coefficient of variation. Computers & Industrial Engineering, 109, 84-95.