Article published online: 2025-01-09

Basic Statistics for Radiologists: Part 1—Basic Data Interpretation and Inferential Statistics

Adarsh Anil Kumar[1] Jineesh Valakkada[1] Anoop Ayyappan[1] Santhosh Kannath[1]

[1] Department of Imaging Sciences and Interventional Radiology, Sree Chitra Institute of Medical Sciences, Trivandrum, Kerala, India

Indian J Radiol Imaging 2025;35(Suppl S1):S58–S73.

Address for correspondence Jineesh Valakkada, MD, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, Kerala 695001, India (e-mail: jineesh174@gmail.com).

Abstract

A systematic approach to statistical analysis is essential for accurate data interpretation and informed decision-making in the rapidly evolving field of radiology. This review provides a comprehensive overview of the fundamental statistical concepts for radiologists and clinicians. The first part of this series introduces foundational elements such as data types, distributions, descriptive and inferential statistics, hypothesis testing, and sampling methods. These are crucial for understanding the underlying structure of research data. The second part of this series delves deeper into advanced topics, including correlation and causality, regression analysis, survival curves, and the analysis of diagnostic tests using contingency tables and receiver operator characteristic (ROC) curves. These tools are vital for evaluating the efficacy of imaging techniques and drawing valid conclusions from clinical studies. As radiology continues to push the boundaries of technology and therapeutic interventions, mastering these statistical principles will empower radiologists to critically assess literature, conduct rigorous research, and contribute to evidence-based practices. Despite the pivotal role of statistics in radiology, formal training in these methodologies is still limited to a certain extent. This primer aims to bridge that gap, providing radiologists with the necessary tools to enhance diagnostic accuracy, optimize patient outcomes, and advance the field through robust research.

Keywords

- central tendency
- hypothesis testing
- parametric
- statistics

Introduction

Adopting a systematic approach to statistical analysis is essential for ensuring the accurate interpretation of data and drawing valid conclusions from research studies. In the field of radiology, statistics play a crucial role in enhancing diagnostic precision, improving patient outcomes, and driving advancements in research. This primer offers a thorough and condensed overview of key statistical concepts that are pertinent to both radiologists and clinicians. The first part is dedicated to discussing types of data, data distribution, descriptive and inferential statistics, hypothesis testing, and sampling. The second part delves into advanced statistical concepts such as correlation and causality, regression analysis, survival curves, and the analysis of diagnostic tests, encompassing contingency tables and receiver operating characteristic (ROC) curves. This primer not only serves as a foundational resource for grasping basic statistical concepts but also aids in the interpretation of various methodologies relevant to daily research endeavors.

Radiology has been at the forefront of technological innovations and various advancements, focusing not only on disease diagnosis but also on therapeutic interventions. The conduct of research assessing the utility of imaging techniques and their applications are crucial for shaping clinical recommendations and establishing practice guidelines, both now and in the future.[1] Understanding

DOI https://doi.org/10.1055/s-0044-1796644.
ISSN 0971-3026.

Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

fundamental statistical principles will enable radiologists as well as clinicians to critically assess existing literature and make well-informed clinical decisions, which are the foundations of evidence-based medicine.[2] Similarly, the proper application and interpretation of statistical methods are crucial for carrying out scientifically rigorous studies. Nonetheless, training in research methodology, particularly in statistics, is generally limited throughout postgraduate medical training.[3] Our objective is to provide an overview of the most frequently used data analysis methods found in radiology literature.

Types of Data

Statistical data can be broadly classified into two types: quantitative and qualitative. Understanding the type of data are crucial for selecting the appropriate statistical method for analysis.[4] Quantitative data refers to numerical information that can be measured and counted. It can be further subdivided into two types (►**Fig. 1**)[5,6]:

- *Continuous data* can take any value within a specified range, allowing for the calculation of statistical measures such as means and variances. For instance, in a study measuring the size of tumors in breast cancer patients before and after treatment, the tumor sizes are considered continuous data because they can assume any value within the range of possible measurements, such as 1.2, 2.5, 3.7 cm, and so on.
- On the other hand, *discrete data* consist of distinct and separate values, often arising from counting processes. For example, the number of renal cysts present on ultrasound images of different patients represents discrete data. If one patient has three cysts and another has five, these values are discrete data.

Qualitative data describe characteristics or categories that cannot be quantified. They are also known as categorical data and can be subdivided into two types[5,6]:

- *Nominal data:* These represent categories that do not have an inherent order. This type of data is often used to classify observations into distinct groups. For example, in a study evaluating the choice of different imaging modalities for a particular suspected pathology among various radiologists, the modalities (magnetic resonance imaging [MRI], computed tomography [CT], ultrasound) are nominal data.
- *Ordinal data:* This type of data represents categories with a meaningful order but no consistent difference among them. It is useful for ranking observations but does not provide information about the relative distance between ranks. For example, when evaluating patient satisfaction with imaging services, responses might be categorized as "poor," "fair," "good," or "excellent." These categories have a natural order, but the intervals between them are not necessarily equal.

Consider a study that examines the efficiency of different radiology workflows. The study can collect both quantitative and qualitative data. *Quantitative data* can be measured as the time taken (in minutes) to complete a set of imaging examinations, while *qualitative data* can be formulated as the type of workflow (manual vs. automated). Statistical tests are

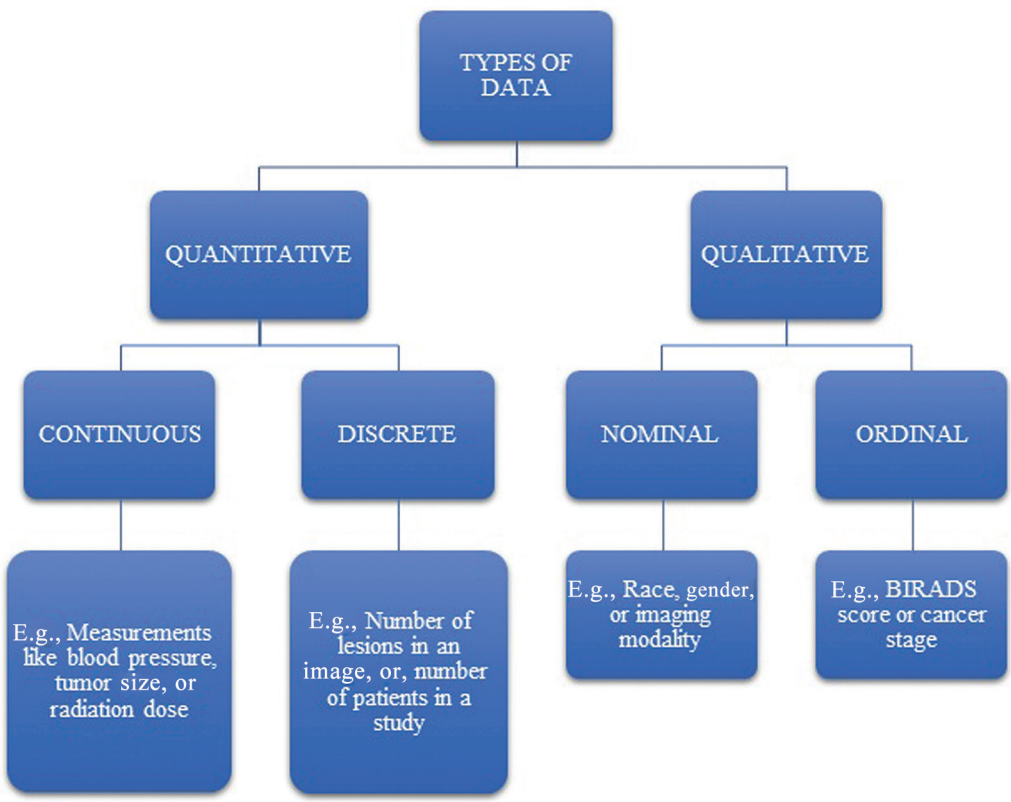

Fig. 1 Flowchart demonstrating the classification of types of data.

more robust for quantitative data than for qualitative data. By analyzing both types of data, the researcher can determine not only which workflow is faster but also how the type of workflow affects overall efficiency as well as user satisfaction.

When gathering data for research, it is advisable to collect the data as continuous variables rather than nominal variables when there is flexibility in organizing the data. For instance, when recording the hypertensive status of multiple patients, it is more advantageous to gather individual blood pressure measurements rather than categorizing patients as hypertensive or nonhypertensive. This approach offers benefits such as greater statistical power, reduced information loss, and increased flexibility in data transformation.

Distribution of Data

Understanding the distribution of data is essential for selecting appropriate statistical methods. Distribution describes how the data values are spread across and thereby provides insight into underlying patterns as well as trends within the dataset.[7]

Normal distribution (also known as Gaussian distribution) basically links frequency distribution to probability distribution, representing how near or how far distribution of the observed sample is from the ideal distribution of a population-based sample. It is a symmetrical, bell-shaped curve where most of the data points cluster around the mean. Many biological measurements, like blood pressure or body temperature, follow a normal distribution. Mean in such data occupies the central position within the distribution. Standard deviation (SD) indicates how data are dispersed around the mean. Larger the SD, wider and flatter the curve. Two SDs cover 95% and 3 SDs cover 99.7% of the observations. The properties of the normal distribution allow for the application of various statistical techniques, including parametric tests.[7,8]

Skewness is a measure of asymmetry and deviation from a normal distribution. Data can be skewed if they are not symmetrically distributed. Skewness can be positive (right skewed) or negative (left skewed; ►**Fig. 2**).[9]

Right-skewed distribution: Most data points are concentrated on the left with a long tail to the right. For example, in a dataset measuring the duration of hospital stays for patients

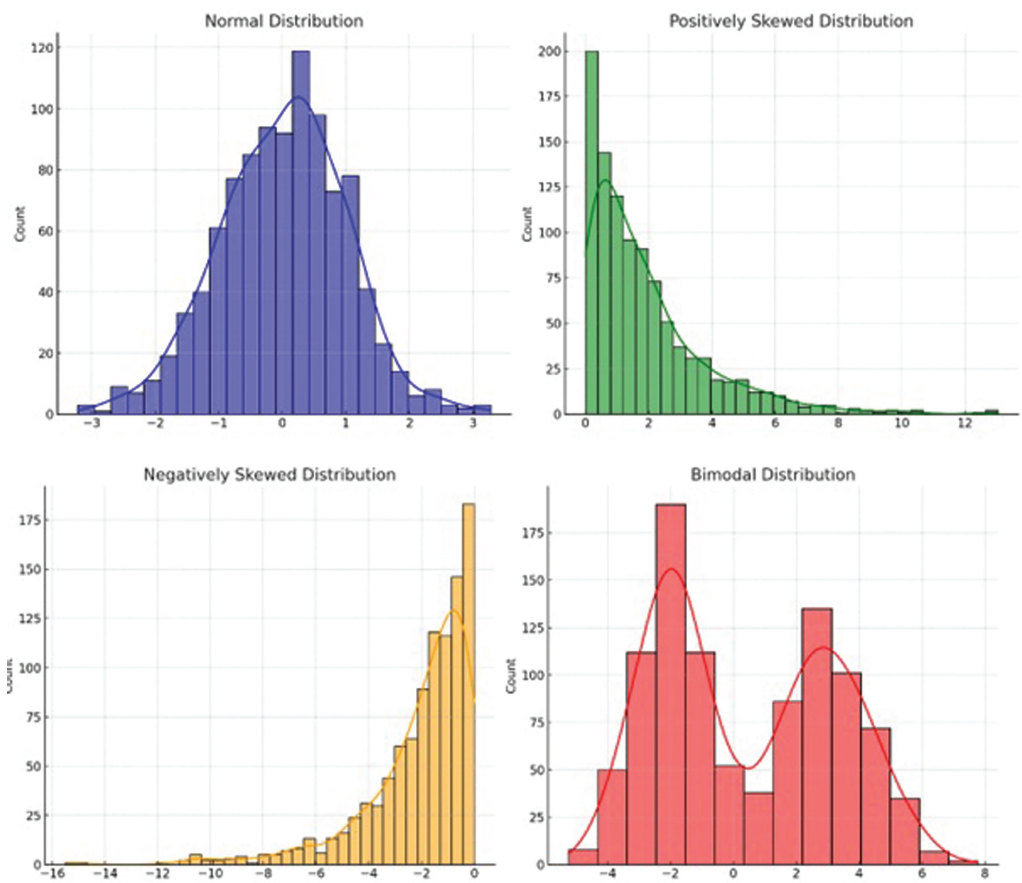

Fig. 2 Bar charts demonstrating types of data distribution. Normal distribution of data is represented by the typical symmetrical bell-shaped curve, e.g., in a typical healthy population, liver attenuation values (in HU) usually center around a mean of 50 to 60 HU, with most people falling close to this value. There are few individuals with extremely high or low attenuation values, leading to the characteristic bell-shaped, symmetrical curve of a normal distribution. Positively skewed distribution causes the peak of the curve to shift toward the positive left side, e.g., in a dataset measuring duration of hospital stays for patients undergoing different interventional radiology procedures, a right-skewed distribution might indicate that while most patients are discharged within a few days, a smaller number of patients have significantly longer stays due to complications. Negatively skewed distribution causes it to shift toward the negative right side, e.g., if age at diagnosis for a particular disease shows a left-skewed distribution, it might indicate that most diagnoses occur later in life, with a few cases occurring at younger ages. Bimodal distribution with two peaks on the right and left side, for example, distribution of heights in a mixed-gender sample.

undergoing different interventional radiology procedures, a right-skewed distribution might indicate that while most patients are discharged within a few days, a smaller number of patients have significantly longer stays due to complications.

Left-skewed distribution: Most data points are concentrated on the right with a long tail to the left, such as in the case of age at diagnosis for a particular disease. For example, if age at diagnosis for a particular disease shows a left-skewed distribution, it might indicate that most diagnoses occur later in life, with a few cases occurring at younger ages.

A bimodal distribution has two peaks. This can occur when data are collected from two different populations. For example, the distribution of heights in a mixed-gender sample.

Presentation of Data

Data can be presented in three ways: as text, in tabular form, or in graphical form (►**Fig. 3**)[4,10]:

- *Text*: This is the main method of conveying information to explain results and trends, as well as to provide contextual information.
- *Table*: It helps in the representation of larger amounts of data in an engaging, easy-to-read and coordinated manner. The data are arranged in rows and columns.
- *Graphical form*: It is a powerful tool to communicate research results and to gain information from data. It may be in the form of a bar chart, pie chart, line diagram, scatter plot, or histogram.

Descriptive and Inferential Statistics

Once you have gathered data and organized it according to its type and distribution, the next step is to analyze the data. One important aspect of statistics involves making assertions about a population. Since it is often impractical to obtain data from an entire population, a sample is typically taken instead. Descriptive statistics are then used to characterize this sample, including measures such as the mean value and the degree of dispersion. However, characterizing the sample alone does not provide insight into the population as a whole; this is the domain of inferential statistics. In this case, a sample is drawn from the population with the aim of drawing broader conclusions about the population based on

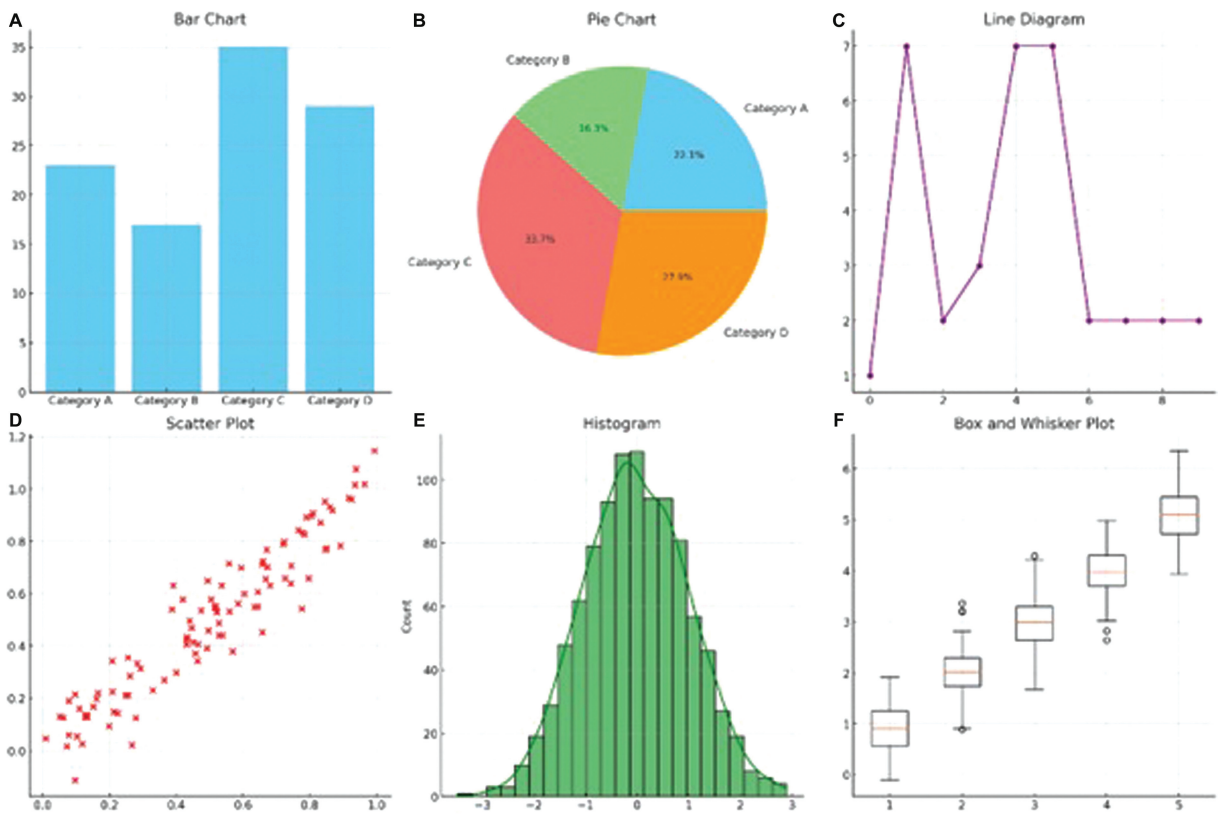

Fig. 3 Examples of different forms of data presentation. (**A**) Bar chart, which is used to compare the frequency or values of different categories, for example, comparing the number of patients with different types of brain tumors [gliomas, meningiomas, metastases] diagnosed over a year. (**B**) Pie chart, which is used to show proportions or percentages of a whole, for example, showing the percentage distribution of different imaging modalities (magnetic resonance imaging [MRI], computed tomography [CT], ultrasound, X-ray) used in a hospital's radiology department. (**C**) Line diagram, which is used to track changes or trends over time, for example, tracking the trend of average radiation dose per CT scan in a radiology department over time (across months or years). (**D**) Scatter plot, which is used to explore relationships or correlations between two continuous variables, for example, plotting the relationship between tumor size (in cm) and patient survival time (in months) after diagnosis of a malignant tumor. (**E**) Histogram, which is sed to display the distribution of a continuous variable by grouping data into bins, for example, displaying the distribution of radiodensity values (in Hounsfield units) for liver tissue on CT in a group of patients to assess for fatty liver disease. (**F**) Box and whisker plot, which is used to show the spread, central tendency, and outliers in a dataset, for example, comparing the distribution of radiologists' interpretation times (in minutes) for reading brain MRI across different experience levels (junior, senior, expert).

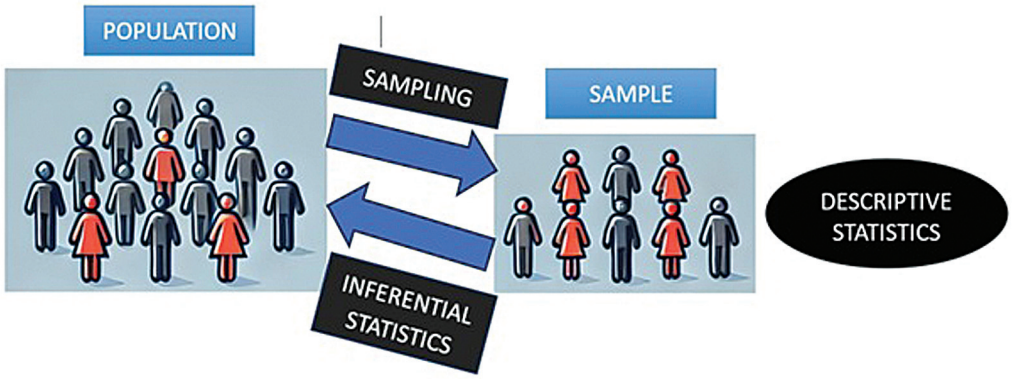

Fig. 4 Pictorial representation of descriptive versus inferential statistics. Sampling is the process of selecting a subset of individuals or data points from a population to make inferences about the entire population. Inferential statistics are used to make predictions or generalizations about a population based on sample data, often involving hypothesis testing and confidence intervals. Descriptive statistics are used to summarize and describe the main features of a dataset, such as measures of central tendency and variability.

this sample. Thus, inferential statistics seek to deduce the unknown parameters of the population from the known parameters of a sample, going beyond the immediate data unlike descriptive statistics. To accomplish this, inferential statistics utilize hypothesis tests such as the *t*-test or analysis of variance (ANOVA). Both are crucial for analyzing data and drawing meaningful conclusions from them (►**Fig. 4**).[11]

Descriptive Statistics

Descriptive statistics summarize and describe features of a particular dataset using statistical characteristics, graphics, charts, or tables. They provide simple summaries about the sample and its measures, thereby offering critical insights into central tendency, dispersion, and shape of data distribution. It is important to understand that in descriptive statistics only properties of the sample are evaluated, and we do not draw conclusion about other points in time or the population. Descriptive statistics are further broadly divided into two subtypes: location parameters (i.e., measures of central tendency) and dispersion parameters (i.e., measures of variability). Parameter basically represents a measurable characteristic of the population.

Measures of Central Tendency

Measures of central tendency basically describe where the center of a sample is or where most of the sample is.[12–14]

Mean: it represents the average of all data points, which is calculated by summing all the values and dividing by the number of observations. The mean can be calculated only for metric variables and is sensitive to outliers. For example, if a radiologist measures the mean size of the liver in a sample of five patients with glycogen storage disorders as 15, 16, 17, 18, and 19 cm, the mean liver size is $(15+16+17+18+19)/5=17$ cm.

Median: when data points are ordered from smallest to largest, the middle value is termed as median. The variables must have an ordinal or metric scale level for calculating median. The median is less affected by outliers and skewed data. For the aforementioned example of liver size in a sample of five patients with glycogen storage disorders, the median is 17. For an even number of observations, the median is the average of the two middle values.

Mode: the most frequently occurring value in the dataset is defined as mode. There can be more than one mode if multiple values have the same frequency. It can be used for metric, nominal, or ordinal variables. For example, if the liver sizes are 15, 16, 17, 17, and 18 cm, the mode is 17 cm because it appears most frequently. The advantages and disadvantages of measures of central tendency are given in ►**Table 1**.

Measures of Variability

Measures of variability describe how much values of variables in a sample differ from each other. In other words, they described how much the values of the variable deviated from the mean value (►**Fig. 5**).[15–18]

Range: it is the difference between the highest and lowest values in the dataset. It gives a sense of the spread but is affected by outliers. Let us consider the previous example of a radiologist measuring the mean size of the liver in a sample of five patients with glycogen storage disorders as 15, 16, 17, 18, and 19 cm. Range is $19–15=4$.

Variance: the average of the squared differences from the mean. Variance provides a measure of how much the values in the dataset deviate from the mean.

For a population, the formula is the following:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

where N is the size of the population; x_i are the values in the population, μ is the population mean.

For a sample, the formula is the following:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

where n is the size of the sample, x_i are the values in the sample, $\bar{x}$ is the sample mean.

Table 1 Table demonstrating the advantages and disadvantages of measures of central tendency

Measure of central tendency	Advantages	Disadvantages
Mean	• Takes all data points into account, providing a comprehensive summary	• Sensitive to outliers, which can skew the result
	• Most commonly used and understood	• Not suitable for skewed distributions
Median	• Not affected by outliers or skewed data	• Does not consider all data points, only the middle value
	• Represents the 50th percentile, providing a central location	• Less informative in symmetric distributions with no outliers
Mode	• Useful for categorical data where we wish to know the most common category	• May not be unique or may not exist in a continuous dataset
	• Not affected by outliers	• Less informative when the distribution is fairly uniform

For the example mentioned above (liver sizes of 15, 16, 17, 18, and 19 cm), the variance is calculated as the following:

- Calculate the mean: $\bar{x} = (15 + 16 + 17 + 18 + 19)/5 = 17$.
- Calculate the squared differences from the mean: $(x_i - \bar{x})^2$–
  - $(15 - 17)^2 = (-2)^2 = 4$.
  - $(16 - 17)^2 = (-1)^2 = 1$.
  - $(17 - 17)^2 = 0^2 = 0$.

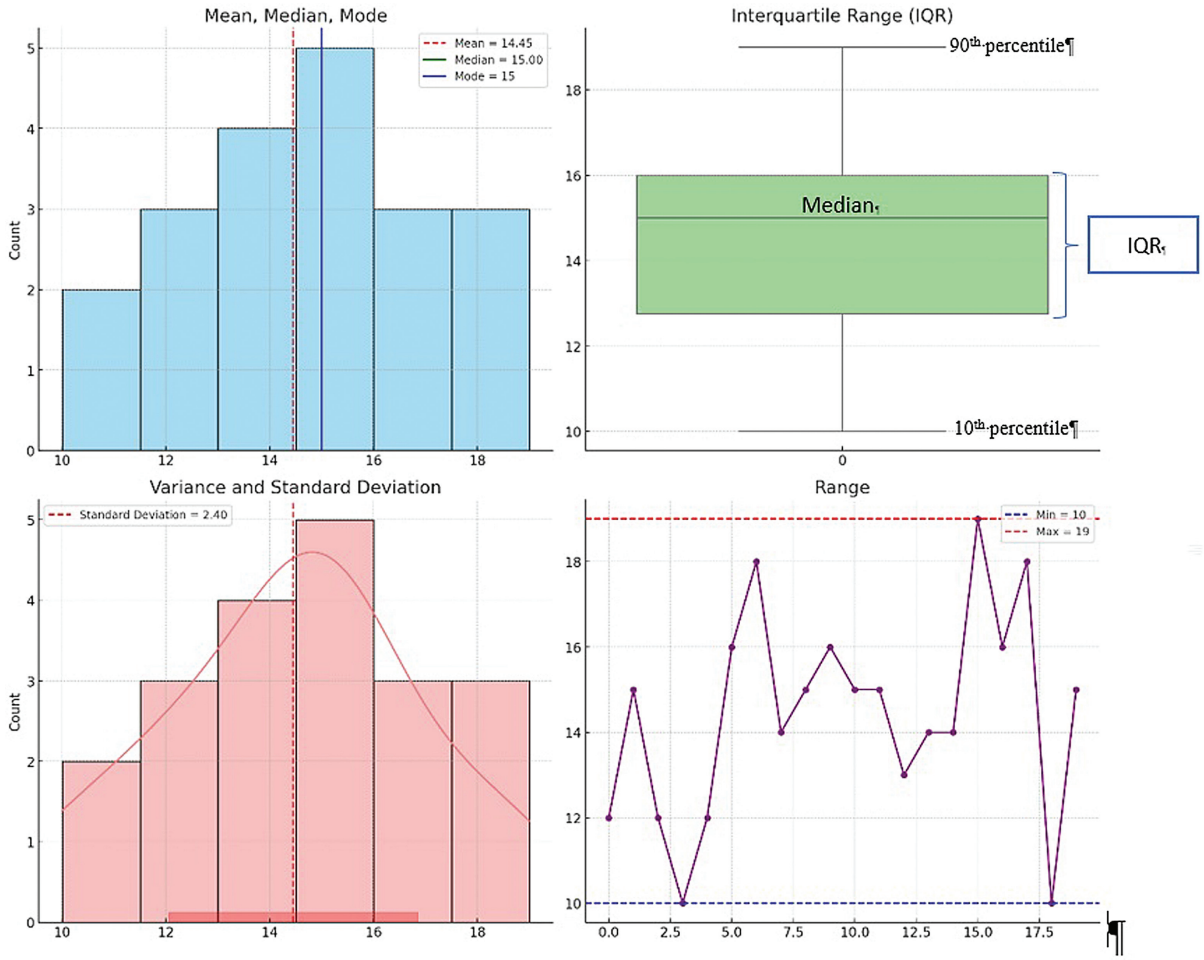

Fig. 5 Graphical representation of measures of central tendency and measures of dispersion. Measures of central tendency are statistical metrics (mean, median, mode) that represent the central point or typical value in a dataset, for example, if a radiologist measures the mean size of the liver in a sample of five patients with glycogen storage disorders as 12, 15, 15, 16, and 14 cm, the mean liver size is $(12 + 15 + 15 + 16 + 14)/5 = 14.5$, the median is 15, and, mode is 15. Measures of dispersion on the other hand are metrics (range, variance, standard deviation) that quantify the spread or variability of data around the central tendency, for example, in the previous example of mean liver size measurement, if the values are 10, 13, 14, 16, and 19 cm, range will be 9, variance will be 9.31, and standard deviation will be 3.05.

- $(18-17)^2 = 1^2 = 1$.
- $(19-17)^2 = 2^2 = 4$.

- Sum the squared differences: $\sum n(x_i - x)^2 = 4 + 1 + 0 + 1 + 4 = 10$.
- Calculate the variance: $s^2 = 10/(5-1) = 10/4 = 2.5$.

SD: it is the square root of variance and indicates the average distance of data points from the mean. Thus, SD is the mean deviation (root mean square) of all measured values from the mean. It is expressed in the same units as the data.

For a population, the formula is the following:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

where N is the size of the population, x_i are the values in the population, and μ is the population mean.

For a sample, the formula is the following:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where n is the size of the sample, x_i are the values in the sample, and $\bar{x}$ is the sample mean.

For the example mentioned above (liver sizes of 15, 16, 17, 18, and 19 cm), SD is calculated as the following:

- Calculate the variance: $s^2 = 2.5$.
- Calculate the SD: $s = \sqrt{s^2} = \sqrt{2.5} = 1.58$.

Quartile: it divides data into four parts as equal as possible. For this, the data must be arranged from the smallest to the largest.

- Quartile (Q1): Middle value between the smallest value and the median.
- Quartile (Q2): Median of the data, that is, 50% of the values are smaller and 50% of the values are larger.
- Quartile (Q3): Middle value between the median value and the largest value.

Interquartile range: to find out the range in which the middle 50% of all values lie, one can use the scattering parameter known as interquartile range.

The advantages and disadvantages of measures of variability are given in ►**Table 2**.

Inferential Statistics

Inferential statistics allow us to make predictions or inferences about a specific population based on the sample data. This includes estimating population parameters as well as testing hypotheses. It therein provides a way to generalize findings beyond the observed data.[19]

Inferential statistics are broadly of four types:

- Difference between two groups of variables.
- Correlation between two groups of variables.
- Predicting the outcome variable.
- Relation of variables in time distribution.

In this section, we shall be dealing with the difference between two groups of variables. The rest will be dealt with in part 2 of the series.

Estimation

Estimation refers to the use of sample data to estimate population parameters, such as the mean or proportion. The accuracy of these estimates can be assessed using confidence intervals.[20]

Confidence intervals: range of values within which the true population parameter is expected to lie with a certain level of confidence (e.g., 95% confidence interval). A wider interval indicates greater uncertainty about the parameter estimate. Let us consider the example of a study measuring the average radiation dose patients receive during a whole

Table 2 Table demonstrating the advantages and disadvantages of measures of variability

Measure of central tendency	Advantages	Disadvantages
Range	• Simple and easy to calculate	• Highly sensitive to outliers
		• Ignores the distribution of data points within the range
Variance	• Takes into account all data points, providing a comprehensive measure	• Not in the same units as the original data (squared units)
	• Useful in statistical calculations and inferential statistics	• Sensitive to outliers
Standard deviation	• Provides a clear measure of spread in the same units as the original data	• Sensitive to outliers
	• Widely used and understood in statistical analysis	• Can be less intuitive to interpret compared to the range
Interquartile range	• Not affected by outliers, as it focuses on the middle 50% of data	• Ignores the data outside the 1st and 3rd quartiles
	• Useful in skewed distributions	• Less informative for distributions that are not skewed or have outliers

body 18-FDG positron emission tomography (PET)/CT, where a 95% confidence interval might be 13 to 15 mSv. The confidence level of 95% means that if we were to repeat this study multiple times, approximately 95% of the calculated confidence intervals from those studies would contain the true population mean radiation dose.

Hypothesis Testing: Fundamentals

Hypothesis is defined as an assumption that is neither proved nor disproved. It is a research process that involves testing assumptions or claims about a population parameter. Usually hypotheses are formulated starting from a literature review and framing a research question based on this review. Hypothesis testing of the collected data provides a formal framework for making decisions based on sample data. The final target is to either reject or retain this hypothesis.[21,22]

Null and Alternative Hypothesis

Null hypothesis (H0): it is the default assumption that there is no statistically significant difference between two or more groups with respect to a particular characteristic (like no statistically significant difference between variables or no effect of an intervention). In a study comparing two imaging techniques, the null hypothesis might state that there is no statistically significant difference in the diagnostic accuracy between these two techniques.

Alternative hypothesis (H1): alternate hypothesis assumes that there is a difference between two or more groups. It represents the opposite of the null hypothesis. Alternative hypothesis might state that there is a difference in diagnostic accuracy between the two imaging techniques.

Difference and Correlation Hypothesis

Difference hypothesis: it tests whether there is a difference between two or more groups. Difference hypothesis might state that there is a difference in diagnostic accuracy between two imaging techniques.

Correlation hypothesis: it tests whether there is a correlation between two or more variables. Correlation hypothesis might state that there is a correlation between the size of a tumor measured by ultrasound and its volume measured by MRI.

Directional and unidirectional hypothesis: with an undirectional hypothesis, focus of interest is whether there is a difference in a value between the groups under consideration. On the other hand, a directional hypothesis focuses on whether one group has a higher or lower value than the other.

The fundamental concept of hypothesis testing is that whether a hypothesis can be accepted or rejected based on a certain probability of error. The reason for this probability of error is that each time you take a sample, you get a different sample, which means that the results are different every time.[23]

Type I error: it refers to rejecting the null hypothesis when it is true (false positive). The significance level (α) represents the probability of making a type I error. Usually, a significance level of 5 or 1% is set.

For example, if α is set at 0.05, there is a 5% chance of incorrectly rejecting the null hypothesis when it is actually true.

p-Value: it is the probability of obtaining the observed results if the null hypothesis is true. If the *p*-value is less than the significance level, the null hypothesis is to be rejected (otherwise not). A *p*-value less than 0.05 is typically considered statistically significant, indicating that the observed results are unlikely to have occurred by chance. For example, if the *p*-value is 0.03 in a study comparing imaging techniques, it suggests that there is a statistically significant difference in diagnostic accuracy.

Type II error: it is failing to reject the null hypothesis when it is false (false negative). The probability of making a type II error is denoted by β, and power is defined as 1–β. For example, if a study has low power, there is a higher chance of failing to detect a true difference between imaging techniques, resulting in a type II error.

It is important to keep in mind that just because an effect is statistically significant it does not mean that the effect is relevant. If a very large sample is taken and it has a very small spread, even a minute difference between two groups may be significant, but it may not be practically relevant.

Sample Size Determination

Determining the appropriate sample size is very crucial for ensuring the reliability and validity of study results. Too small a sample size will not give valid results or will not adequately represent the realities of the population being analyzed. On the other hand, larger sample sizes give smaller margins of error and are more representative. In fact, a sample size that is too large may significantly increase the cost and time taken to conduct the research.[24–28] The factors that influence sample size include the following:

- Population size: larger populations generally require larger samples.
- Effect size: smaller effect sizes require larger samples to detect differences.
- SD: the higher the distribution is, the greater the SD and the greater the magnitude of deviation.
- Significance level (α): lower significance levels require larger samples.
- Power ($1-\beta$): higher power (typically 0.80) requires larger samples to reduce the risk of type II errors.

Case Study: Sample Size in Radiological Research

A study aims to evaluate the diagnostic accuracy of a new MRI sequence in neuroimaging. Researchers need to determine an appropriate sample size to ensure the study's findings are statistically significant and reliable.

- Population size: the population includes all patients eligible for brain MRI at the hospital.
- Effect size: based on preliminary data, the researchers estimate a moderate effect size.
- Significance level (α): they choose a significance level of 0.05.
- Power ($1-\beta$): they aim for a power of 0.80, meaning they want an 80% chance of detecting a true difference if one exists.

Using sample size calculation formulas, they determine that a sample size of 200 patients is needed to achieve the desired power and significance level. This ensures that the study results will be robust and reliable, providing valuable insights into the new MRI technique's diagnostic accuracy.

But which formula should we use to calculate the sample size (►**Fig. 6**, ►**Table 3**)?

Confidence level	*z*-score
80%	1.28
85%	1.44
90%	1.65
95%	1.96
99%	2.58

Steps in using the formula for sample size calculation:

1. Determine the population size (if known).
2. Determine the confidence interval.
3. Determine the confidence level.
4. Determine the SD (basically representing the population proportion, which is assumed to be 50% = 0.05).[29]
5. Convert the confidence level into a *Z*-score.
6. Put these figures into the sample size formula to get your sample size.

Necessary sample size = (Z-score)^2 × SD × (1–SD)/(margin of error)^2.

Say you choose to work with a 95% confidence level, an SD of 0.5, and a confidence interval (margin of error) of ± 5%.

Necessary sample size = {$(1.96)^2 \times 0.5 \times 0.5/(0.5)^2$} = $(3.8416 \times 0.25)/0.0025 = 384.16$.

Hence, the sample size should be 385.

Hypothesis Testing

Hypothesis testing is a statistical method used to make decisions about the population based on sample data. It is used to assess whether a particular viewpoint is likely to be true.[30] It involves several steps (►**Fig. 7**):

1. Formulate hypotheses: define the null hypothesis (H0) and alternative hypothesis (H1).
2. Selection of study design and sample size: select ones that are appropriate to the hypothesis being tested.
3. Select significance level (α): commonly set at 0.05.
4. Collect data: gather sample data relevant to the hypothesis.
5. Calculate test statistic: use an appropriate test (e.g., *t*-test, chi-squared test) to calculate the test statistic for each outcome variable of interest.
6. Determine *p*-value: compare the *p*-value to the significance level.
7. Make a decision: reject H0 if *p*-value $< \alpha$; otherwise, fail to reject H0.

Hypothesis testing is just like the concept of "An accused is presumed to be innocent until proved guilty."

Formula·1·(Unlimited·population):

$$n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\epsilon^2}$$

Formula·2·(Finite·population):

$$n' = \frac{n}{1 + \frac{z^2 \times \hat{p}(1-\hat{p})}{\epsilon^2 N}}$$

Formula·3:

$$n = N \times \frac{\frac{Z^2 \times p \times (1-p)}{e^2}}{N - 1 + \frac{Z^2 \times p \times (1-p)}{e^2}}$$

Formula·4:

$$N = \frac{2\sigma^2 (z_{1-\alpha} + z_{1-\beta})^2}{d_{min}^2}$$

Fig. 6 Formulae for sample size. In Eq. 1, *n*: required sample size for an unlimited population; *z*: *Z*-score, corresponding to the desired confidence level (e.g., 1.96 for 95% confidence); $\hat{p}$: estimated proportion of the population (i.e., the proportion you expect to observe a certain characteristic in the population); ε: margin of error (the maximum acceptable difference between the true population parameter and the sample estimate). In Eq. 2, *n'*: adjusted sample size for a finite population; *n*: sample size calculated for an unlimited population (from the first formula); *N*: size of the finite population. In Eq. 3, *n*: required sample size for a finite population; (*N*): total population size; (*Z*): *Z*-score corresponding to the desired confidence level (e.g., 1.96 for 95% confidence); *p*: estimated proportion of the population (the probability of the characteristic being studied); 1–*p*: complementary proportion (the probability of not having the characteristic being studied); *e*: margin of error (acceptable level of precision in the results). In Formula 4, (*N*): required sample size; σ^2: population variance (or an estimate of the variance of the outcome); $Z_{1-\alpha}$: *Z*-score corresponding to the desired level of statistical significance (e.g., 1.96 for a 95% confidence level), which accounts for type I error (false positives); $Z_{1-\beta}$: *Z*-score corresponding to the desired statistical power, representing type II error (false negatives); typically, $1-\beta$ is set at 0.80 or 0.90, and the corresponding *Z*-score is looked up (e.g., 0.842 for 80% power); d_{min}: minimum detectable difference or effect size, representing the smallest difference that is practically significant and you wish to detect in your study.

Table 3 Table showing minimum sample size calculation of different statistical tests and examples with radiology literature citations

Test type	Formula	Variables needed	Example in radiology	Study
Unpaired *t*-test	$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{(M_1 - M_2)^2}$	• Significance level (α) • $Z_{\alpha/2}$ is the *Z*-value corresponding to the desired significance level • Power (1–β) • $Z_{1-\beta}$ is the *Z*-value corresponding to the desired power • Standard deviation (σ) • Effect size (difference in means; M1–M2)	Comparison of 320-detector volumetric and 64-detector helical computed tomography (CT) images of the pancreas for size measurement of various anatomical structures	Goshima et al[48]
Paired *t*-test	$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \sigma_d^2}{d^2}$	• Significance level (α) • Power (1–β) • Effect size (mean difference *d*) • Standard deviation of differences (σd)	Comparison of tumor size on microscopy, CT, and MRI assessments vs. pathologic gross specimen analysis of pancreatic neuroendocrine tumors	Bian et al[49]
Chi-squared test	$n = \frac{(Z_{\alpha/2}^2 \cdot p(1-p))}{\Delta^2}$	• Significance level (α) • Proportion (*p*) • Difference in proportions (Δ)	Comparison of enhancement patterns between benign and malignant solid renal lesions	Millet et al[50]
ANOVA	$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \sigma^2}{\eta^2}$	• Significance level (α) • Power (1–β) • Effect size (η^2) • Variance between groups (σ^2)	Population-stratified analysis of bone mineral density distribution in cervical and lumbar vertebrae of Chinese from quantitative computed tomography	Zhang et al[51]

Common Hypothesis Tests in Radiology

It is broadly divided into two groups: hypothesis tests done on numerical data and those done on categorical data. Basically, these tests are used to find the difference between two groups of variables.

Datasets will have to be treated as paired if they are related. Thus, if we compare the systolic blood pressure values of two independent sets of subjects, it is an example of unpaired data. However, if a condition is included like all the individuals in one dataset are siblings of the individuals represented in the other dataset, then corresponding values in the two datasets may be related in some manner (due to genetic or familial reasons) and the datasets are no longer independent.

Parametric data are normally distributed numerical data that follows the parameters of a normal distribution curve. If it is a skewed distribution, there is no particular distribution, or if the distribution is unknown, then it should be considered as nonparametric data. But practically, how do we determine whether the numeric data are normally distributed? One gross method is to look at the measures of central tendency, mean, and median. If the mean and median are the same or are very close to one another (as compared with the total data spread), then we can assume that we are dealing with parametric data. However, the proper method to test the fit of data to a normal distribution is to use "goodness-of-fit" tests such as the Kolmogorov–Smirnov test and Shapiro–Wilk test. The null hypothesis in these tests is that the frequency distribution of your data is normally distributed. If any of these tests return a *p*-value less than 0.05, it implies that the normal distribution will have to be rejected and the data would have to be taken as nonparametric.[31–34]

Statistical tests for normal distribution:

- Kolmogorov–Smirnov test.
- Shapiro–Wilk test.
- Anderson–Darling test.
- D'Agostino–Pearson omnibus test.

The major disadvantage of these tests is that the calculated *p*-value is affected by the sample size. Therefore, if the sample size is very small, the *p*-value may be much larger than 0.05. But if the sample size from the same population is very large, your *p*-value may be smaller than 0.05.

To overcome this disadvantage, graphical tests for normal distribution are used (►**Fig. 8**):

- Histogram data: Compare the histogram curve with the normal distribution curve.
- Quantile–quantile plot: Compare the theoretical quantiles of normally distributed data with quantiles of the measured values. If data were perfectly normally distributed, all the points would be on a straight line. The further the points deviate from the line, the less normally distributed the data are.

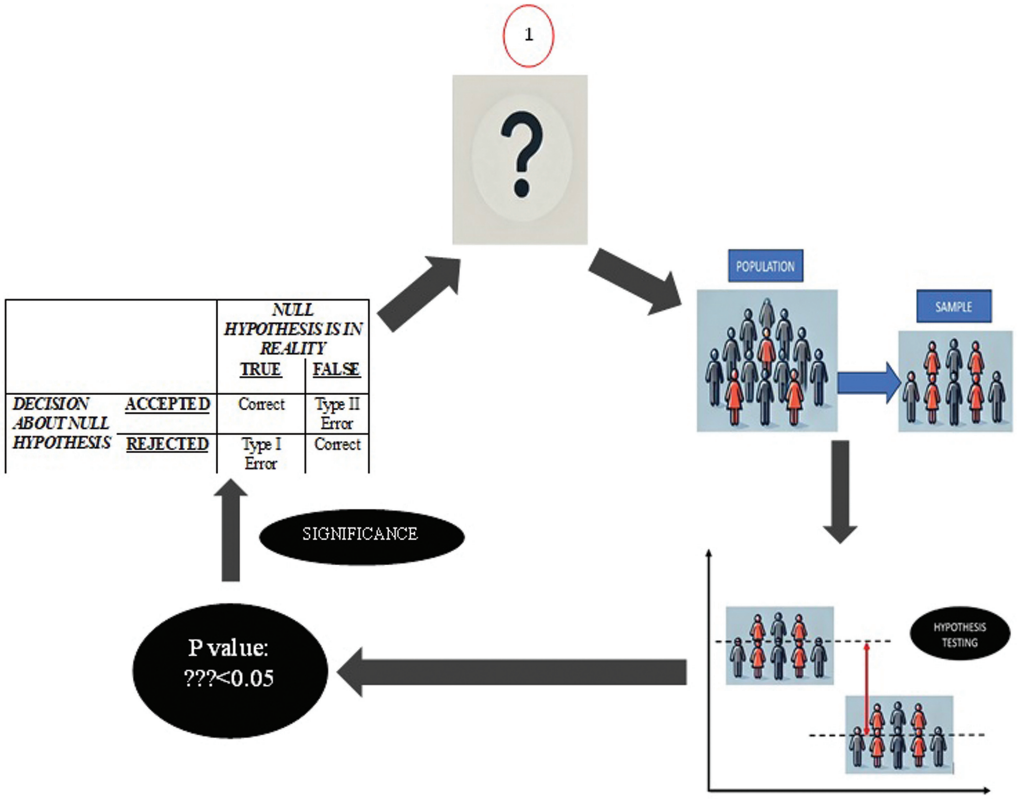

Fig. 7 Pictorial representation of hypothesis testing process. Steps involved in the hypothesis testing process are the following: (1) Formulation of a hypothesis (question mark at the top center). This step involves defining a research question or hypothesis. Typically, there are two hypotheses: (a) null hypothesis (H_0)—assumes no effect or no statistically significant difference and (b) alternative hypothesis (H_1)—assumes there is an effect or difference. (2) Selecting a sample (right panel showing population and sample). From the larger population, a sample is selected. (The sample should be representative of the population to generalize the findings back to the population.) (3) Hypothesis testing (bottom-right panel showing hypothesis testing). Statistical tests are performed on the sample data to test the hypothesis. (The aim is to determine whether the data provide enough evidence to reject the null hypothesis in favor of the alternative hypothesis). (4) Significance and *p*-value (bottom left with *p*-value). The result of the hypothesis test is evaluated using the *p*-value. (If the *p*-value is less than 0.05 [commonly used significance level], it suggests that the results are statistically significant, meaning there is sufficient evidence to reject the null hypothesis.) (5) Conclusion (*Arrow back to the top* indicating significance). Based on the *p*-value and test results, conclusions are drawn about the hypothesis, indicating whether the evidence supports rejecting the null hypothesis.

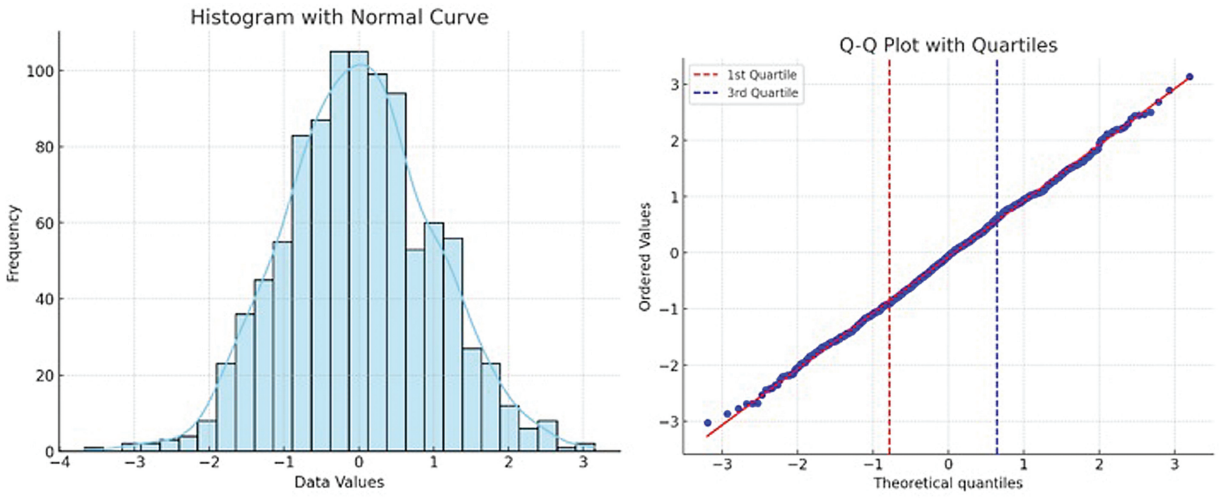

Fig. 8 Histogram curve and Q-Q plot for graphical representation of normality of distribution. Histogram shows the data's shape, and the Q-Q plot compares the data's quantiles to a theoretical normal distribution to identify deviations from normality.

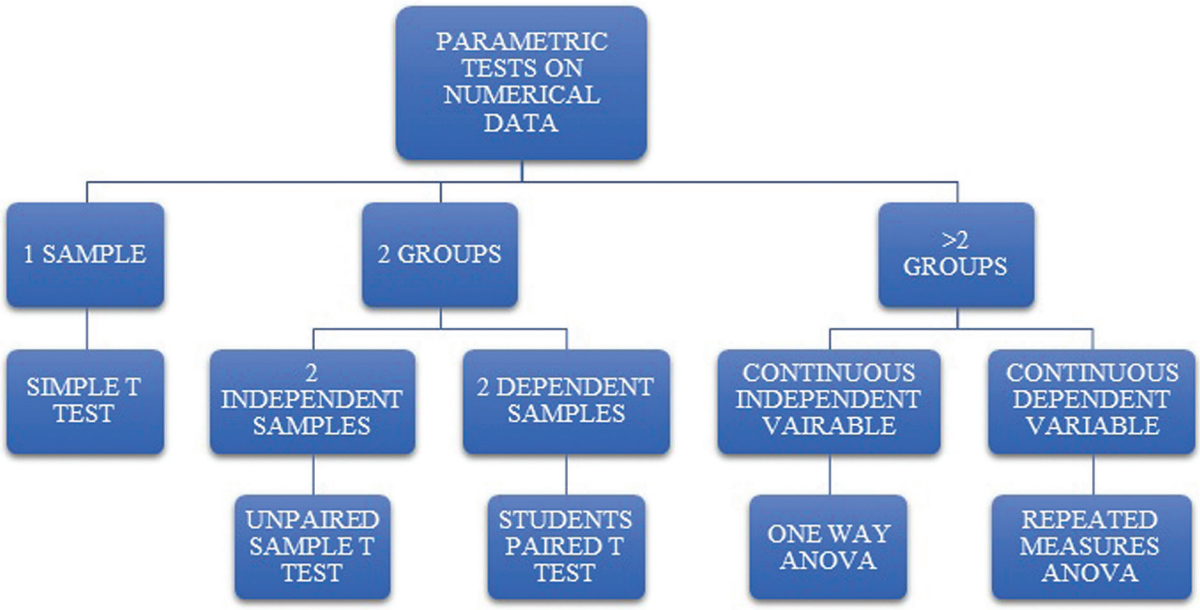

Fig. 9 Approach to select appropriate parametric tests.

Hypothesis Tests Done on Contiguous Data

Parametric Data

Simple *t*-test: this is a test used to determine whether the mean calculated from sample data collected from a single group is different from the population selected (►**Fig. 9**).[35,36]

Let us consider a study where the researchers want to assess whether the hippocampal volume on MRI in temporal lobe epilepsy patients is significantly lower as compared with all epilepsy patients imaged during a specific time period. The *t*-test would then be used to show if the hippocampal volume is statistically lower in temporal lobe epilepsy patients.

Unpaired sample *t*-test (for two independent samples): it compares the means of two independent groups. There is no relationship between the subjects in one group and those in the other.[36] For example, an unpaired *t*-test could be used to compare the average radiation dose received by patients undergoing neurointervention on a monoplane and biplane angio-suite, assuming patients are randomly assigned to one of the techniques.

Student's paired *t*-test (for two dependent samples): it compares the means of two related groups or conditions. Each subject or sample is measured twice, resulting in paired observations.[36] A *t*-test might be used to compare the average size of hepatocellular carcinoma nodules in patients treated with a new intra-arterial chemotherapy drug. If the *t*-test shows a significant difference in mean sizes, it suggests that the drug is effective in reducing tumor size.

A tailed *t*-test refers to either a one-tailed test or a two-tailed test used to determine the direction of an effect, while a nontailed *t*-test typically implies a two-tailed test that assesses for any significant difference without specifying the direction.

One-tailed *t*-test: it tests for the possibility of an effect in one specific direction (e.g., greater than or less than). For example, when the research hypothesis predicts the direction of the difference (e.g., drug A increases recovery rate more than drug B). Basically, it tests if the mean is greater than a certain value.

Two-tailed *t*-test: it tests for the possibility of an effect in both directions (e.g., not equal to). For example, when the research hypothesis does not predict the direction of the difference (e.g., drug A has a different recovery rate than drug B, without specifying higher or lower). Basically, it tests if the mean is different from a certain value, either higher or lower.

One factorial ANOVA (for more than two independent samples): it determines whether there are any statistically significant differences between the means of three or more independent groups (or levels) on a continuous independent variable. It tests the null hypothesis that all group means are equal[37,38]: A one-way factorial ANOVA could be used to compare the average reading times of radiologists interpreting images from three different types of imaging modalities (X-ray, MRI, and CT scan).

Repeated measures ANOVA (for more than two dependent samples): it determines whether there are any statistically significant differences between the means of three or more related groups (or levels) on a continuous dependent variable measured at multiple time points or under different conditions. It accounts for the correlation between measurements taken from the same subject across different conditions or at different time points.[38,39] Repeated measures ANOVA could be used to assess the effectiveness of a new contrast agent in enhancing detection of small cerebral metastatic lesions across multiple time points during an MRI scan session (comparing the detection before contrast administration, immediately after contrast administration and 30 minutes postcontrast administration).

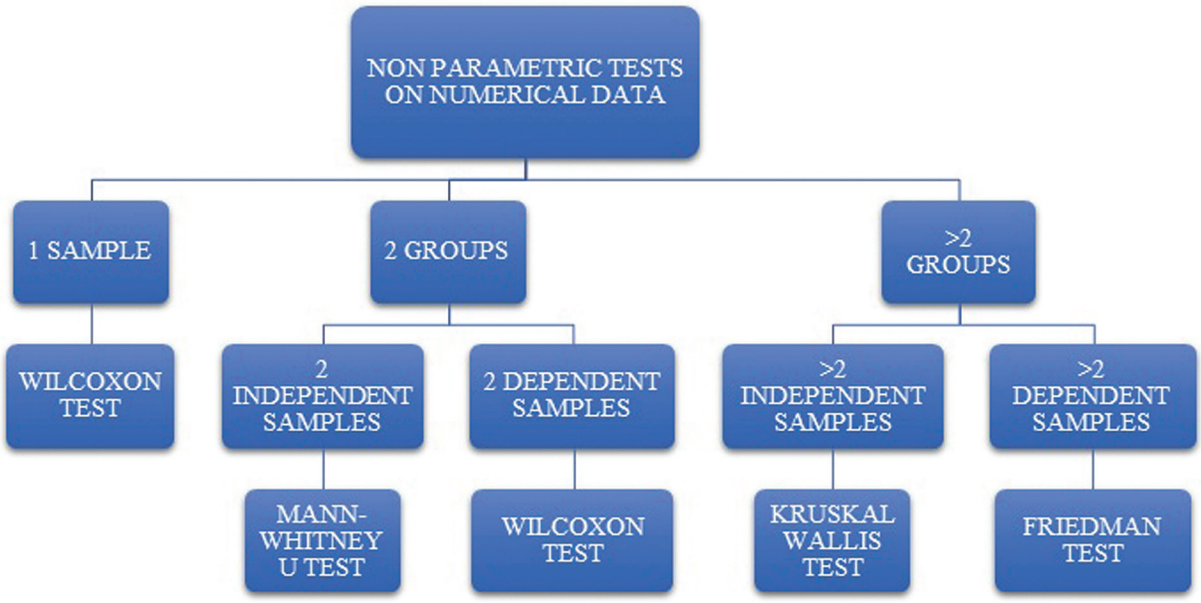

Fig. 10 Approach to select appropriate nonparametric tests.

Non-parametric Data

For One Sample

Wilcoxon's test (Wilcoxon signed-rank test): it compares the median of a single sample of paired data against a specified median value (typically zero, assuming no difference; ►**Fig. 10** and ►**Table 4**). It is typically used when the data do not meet the assumptions required for a parametric test like the *t*-test, such as when the data are not normally distributed or when the measurement scale is ordinal.[40] Wilcoxon signed-rank test could be used to assess whether a new MRI sequence results in significantly improved lesion detection as compared with an established sequence.

Between Two Groups

Mann–Whitney *U* test (for two independent samples; also known as Wilcoxon rank sum test): it assesses whether two independent groups differ significantly in terms of their medians. It does not assume that the data follow a normal distribution.[41] The Mann–Whitney *U* test could be used to compare the interpretation times between two groups of radiologists interpreting the same set of MRI scans.

Wilcoxon's test (for two dependent samples): it compares the medians of two related groups or conditions. It assesses whether there is a statistically significant difference between paired observations from the same subjects under different conditions.[42] Wilcoxon signed-rank test for two dependent samples could be used to evaluate the effectiveness of a new image enhancement AI algorithm compared with the current conventional MRI images.

More than Two Groups

Kruskal–Wallis test (for more than two independent samples): it determines whether there are statistically significant differences between three or more independent groups in terms of their medians. It is an extension of the Mann–Whitney *U* test for more than two groups.[43] For example, the Kruskal–Wallis test could be used to compare the hepatic

Table 4 Table showing various nonparametric tests used depending on the type of variables

Variable type	Test	Description
Continuous	Mann–Whitney *U* test	Compares differences between two independent groups
	Kruskal–Wallis test	Extension of the Mann–Whitney *U* test for three or more groups
Nominal	Chi-squared test	Assesses whether there is a significant association between two categorical variables
	Fisher's exact test	Used for small sample sizes (<5 in 1 cell) to determine if there are nonrandom associations between two categorical variables
	McNemar's test	Used for paired nominal data to determine if there is a difference in proportions
Ordinal	Wilcoxon's test	Compares two independent groups with ordinal data
	Friedman's test	Compares differences between three or more dependent groups (repeated measures) in ordinal scale

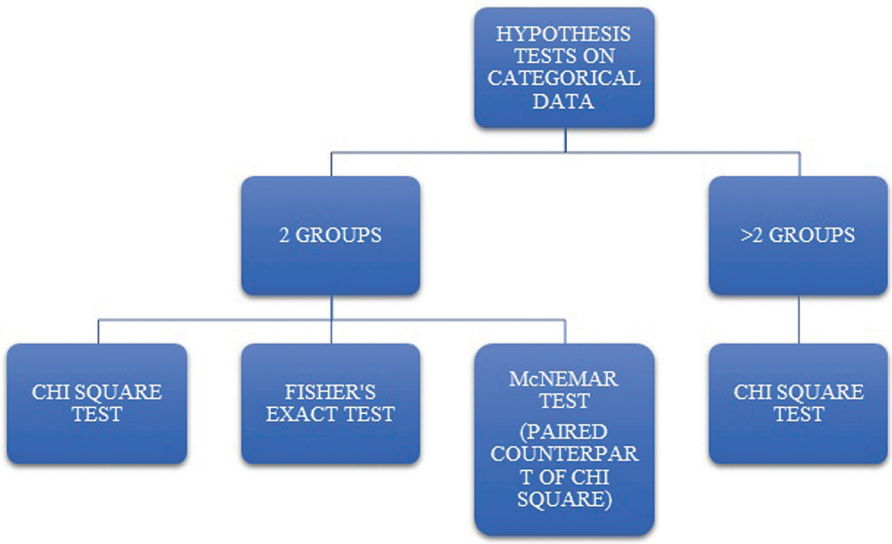

Fig. 11 Approach to select appropriate statistical tests for categorical data.

lesion size (measured as a continuous variable) among three different types of imaging modalities (ultrasound, MRI, and CT scan).

Friedman's test (for more than two dependent samples): it determines whether there are statistically significant differences between three or more dependent groups (repeated measures) in terms of their medians. It is analogous to the Kruskal–Wallis test but is used for within-subject designs.[44] Friedman's test could be used to compare the ratings of definition of margins of a cerebral lesion (ordinal scale) from the same set of radiologists across three different MRI sequences.

Hypothesis Tests Done on Categorical Data

- If two groups are to be compared (►**Fig. 11**)[35]:
  - Chi-squared (χ^2) test: it determines whether there is a significant association between categorical variables. It is typically used when both variables are categorical and the data are frequencies (counts).[45] For example, the chi-squared test could be used to assess the association between the presence of a certain radiological sign and the presence or absence of a specific pathology.
  - Fisher's exact test: it determines whether there is a significant association between categorical variables, especially when sample sizes are small or when expected cell counts in a contingency table are less than 5.[46] For example, Fisher's exact test could be used to compare the diagnostic performance of two imaging techniques in detecting a rare pathology.
  - McNemar's test: it is a nonparametric test used to analyze paired nominal data. It is particularly useful when you have two related samples or repeated measurements on the same subjects, and you want to determine if there is a significant change in responses between two conditions or time points.[47]
- If more than two groups:
  - Chi-squared (χ^2) test: it determines whether there is a significant association between two or more categorical variables. It is an extension of the chi-squared test for two groups but applied to contingency tables with more than two rows or columns.[45] For example, the chi-squared test of independence could be used to assess whether there is an association between the types of lung disease (categorized into four types: pneumonia, tuberculosis, asthma, and bronchitis) and smoking status (smoker vs. nonsmoker) among a group of patients.

The tests to be done based on the type of data are summarized in ►**Tables 4** and **5**.

Table 5 Table showing various parametric and nonparametric tests used depending on the nature of the sample being analyzed

	Parametric tests	Nonparametric tests
One sample	Simple *t*-test	Wilcoxon's test for 1 sample
Two dependent samples	Paired sample *t*-test	Wilcoxon's test
Two independent samples	Unpaired sample t-test	Mann–Whitney *U* test
More than two independent samples	One factorial ANOVA	Kruskal–Wallis test
More than two dependent samples	Repeated measures ANOVA	Friedman's test
Correlation between two variables	Pearson's correlation	Spearman's correlation

Abbreviation: ANOVA, analysis of variance.

Reporting Statistical Tests

Reporting statistical tests in radiology is important to clearly and concisely convey the results of analyses performed to evaluate the significance of findings and robustness of conclusions drawn. Key points to consider when reporting statistical tests are the following:

- *Specify the statistical test used:* clearly mention which statistical test was employed (e.g., *t*-test, ANOVA, chi-squared test, Mann–Whitney *U* test). Justification for the choice of test also has to be provided, including the nature of the data (parametric vs. nonparametric, nominal vs. continuous).
- *Include relevant parameters:* degrees of freedom (if applicable; e.g., for *t*-tests and ANOVA), effect size (include measures such as Cohen's *d* for *t*-tests or eta-squared for ANOVA) to indicate the magnitude of the difference, and confidence intervals (present confidence intervals for mean differences or proportions to give context to the results).
- *Present p-values:* clearly state the *p*-value obtained from the statistical test (use the conventional threshold for significance, e.g., $p < 0.05$; if the *p*-value is above this threshold, avoid stating it as "not significant"; instead, indicate the *p*-value explicitly). For very small *p*-values, it is common to report them as $p < 0.001$.
- *Interpret results:* provide a clear interpretation of what the statistical results mean in the context of the study. Clinical significance of the findings should also be discussed, not just statistical significance.
- *Contextualize with clinical implications:* discuss how the statistical findings relate to clinical practice, patient outcomes, or the diagnostic performance of imaging modalities. Consider including sensitivity, specificity, positive predictive value, and negative predictive value if applicable.
- *Follow reporting guidelines:* adhere to relevant reporting guidelines (e.g., Standards for Reporting Diagnostic Accuracy [STARD] for diagnostic accuracy studies, Consolidated Standards of Reporting Trials [CONSORT] for randomized controlled trials) to ensure clarity and transparency in the reporting of statistical analyses.

Here is an example of how statistical results might be reported in a radiology study.

Let us consider a study to compare the average tumor volume measured by MRI in patients with type A and B tumors. A total of 60 patients were included in the analysis, with 30 patients in the type A group and 30 patients in the type B group. The mean tumor volume for patients with type A tumors was 15.2 cm^3 ($\pm$3.1 cm^3), while the mean tumor volume for patients with type B tumors was 22.8 cm^3 ($\pm$-4.5 cm^3). An independent sample *t*-test was performed to assess whether the difference in mean tumor volumes between the two groups was statistically significant (after testing the normality of distribution).

The results indicated a significant difference in tumor volume between the two groups ($t(58) = -5.46$, $p < 0.001$; "*t*" signifies the result is derived from a *t*-test; the number in brackets is the degree of freedom {N1 + N2–2 = 30 + 30–2 = 58}; –5.46 is the *t* statistic value, with negative indicating the mean of the first group is less than that of the second group; $p < 0.001$ is the *p*-value that is statistically significant). Patients in the type B group exhibited larger tumor volumes than those in the type A group. The effect size, calculated using Cohen's *d*, was 1.41, indicating a large effect. Additionally, a 95% confidence interval for the difference in means was calculated, resulting in an interval of (–9.11 cm^3, –5.25 cm^3). This interval suggests that the mean tumor volume for type B tumors is significantly higher than that for type A tumors, with a clinically relevant difference. In conclusion, these findings demonstrate that patients with type B tumors have significantly larger tumor volumes compared with those with type A tumors, which may have implications for treatment planning and prognosis.

Conclusion

To conclude, statistics play a crucial role in radiology, aiding in accurate data interpretation, improving diagnostic accuracy, and advancing research. Proper understanding and application of statistical principles such as data types, their distribution, descriptive and inferential statistics, hypothesis testing, correlation, and sampling are essential for research in radiology. The foundational knowledge needed to leverage statistics effectively, ultimately enhancing clinical decision-making and patient outcomes.

Authors' Contributions
All the authors were involved in the procedure, data collection, and manuscript revision.

Funding
None.

Conflict of Interest
None declared.

References

1 Psoter KJ, Roudsari BS, Dighe MK, Richardson ML, Katz DS, Bhargava P. Biostatistics primer for the radiologist. AJR Am J Roentgenol 2014;202(04):W365-75
2 Sardanelli F, Hunink MG, Gilbert FJ, Di Leo G, Krestin GP. Evidence-based radiology: why and how? Eur Radiol 2010;20(01):1–15
3 Alderson PO, Bresolin LB, Becker GJ, et al; Consensus Conference Participants. Enhancing research in academic radiology departments: recommendations of the 2003 Consensus Conference. J Am Coll Radiol 2004;1(08):591–596
4 Patel S. Medical statistics series: type of data, presentation of data & summarization of data. Natl J Community Med 2021;12(02): 40–44
5 Seltman HJ. Experimental Design and Analysis. Pittsburgh, PA: Carnegie Mellon University; 2018
6 Hoeks S, Kardys I, Lenzen M, van Domburg R, Boersma E. Tools and techniques: statistics—descriptive statistics. EuroIntervention 2013;9(08):1001–1003
7 Choudhury V, Saluja S. Distribution of data. Curr Med Res Pract 2011;1(05):272–282
8 Kaliyadan F, Kulkarni V. Types of variables, descriptive statistics, and sample size. Indian Dermatol Online J 2019;10(01):82–86
9 Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013;38(01):52–54

10 Karlik SJ. Visualizing radiologic data. AJR Am J Roentgenol 2003; 180(03):607–619
11 Byrne G. A statistical primer: understanding descriptive and inferential statistics. EBLIP 2007;2(01):32–47
12 Mohan S, Su MK. Biostatistics and epidemiology for the toxicologist: measures of central tendency and variability-where is the "middle?" and what is the "spread?" J Med Toxicol 2022;18(03): 235–238
13 Manikandan S. Measures of central tendency: median and mode. J Pharmacol Pharmacother 2011;2(03):214–215
14 Manikandan S. Measures of central tendency: the mean. J Pharmacol Pharmacother 2011;2(02):140–142
15 Salha R. Central tendency and variability measures. Accessed November 11, 2024 at: https://www.researchgate.net/publication/353378173
16 Manikandan S. Measures of dispersion. J Pharmacol Pharmacother 2011;2(04):315–316
17 Ciarleglio A. Measures of variability and precision in statistics: appreciating, untangling and applying concepts. BJPsych Adv 2021;27(02):137–139
18 Cooksey RW. Descriptive statistics for summarising data. In: Illustrating Statistical Procedures: Finding Meaning in Quantitative Data. Singapore: Springer; 2020:61–139
19 Hazra A, Gogtay N. Biostatistics Series Module 2: Overview of Hypothesis Testing. Indian J Dermatol 2016;61(02):137–145
20 Hazra A. Using the confidence interval confidently. J Thorac Dis 2017;9(10):4125–4130
21 Kaur J. Techniques used in hypothesis testing in research methodology: a review. Int J Sci Res 2015;4(05):362–365
22 Goldman Daniel S. The basics of hypothesis tests and their interpretations. Accessed November 28, 2024 at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiR8pqb0v6JAxUdyzgGHY2JOU0QFnoECDQQAQ&url=https%3A%2F%2Fosf.io%2Fu2csn%2Fdownload&usg=AOvVaw3d2X5ynHDpS-dJ6oDL0gk2&opi=89978449
23 Kim HY. Statistical notes for clinical researchers: type I and type II errors in statistical decision. Restor Dent Endod 2015;40(03): 249–252
24 Nnodim J, Onyeze V, Chidozie Nwaokoro J, Ifeanyi Obeagu E, Author C, Johnkennedy N. Sample size determination as an important statistical concept in medical research. Madonna Univ J Med Health Sci 2021;1(02):42–49
25 Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. Gastroenterol Hepatol Bed Bench 2013;6(01):14–17
26 Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013;35 (02):121–126
27 Das S, Mitra K, Mandal M. Sample size calculation: basic principles. Indian J Anaesth 2016;60(09):652–656
28 Eng J. Sample size estimation: how many individuals should be studied? Radiology 2003;227(02):309–313
29 Shete A, Asst PD, Dubewar AP. Sample size calculation in Bio statistics with special reference to unknown population. Int J Innov Res Multidisc Field 2020;6(07):236–238
30 Walker J. Hypothesis tests. BJA Educ 2019;19(07):227–231
31 Altman DG, Bland JM. Statistics notes: the normal distribution. BMJ 1995;310(6975):298
32 Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10(02): 486–489
33 Habibzadeh F. Data distribution: normal or abnormal? J Korean Med Sci 2024;39(03):e35
34 Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 2019;22(01):67–72
35 Anvari A, Halpern EF, Samir AE. Statistics 101 for radiologists. Radiographics 2015;35(06):1789–1801
36 Xu M, Fralick D, Zheng JZ, Wang B, Tu XM, Feng C. The differences and similarities between two-sample t-test and paired t-test. Shanghai Jingshen Yixue 2017;29(03):184–188
37 Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. Application of student's *t*-test, analysis of variance, and covariance. Ann Card Anaesth 2019;22(04):407–411
38 Kim HY. Statistical notes for clinical researchers: a one-way repeated measures ANOVA for data with repeated observations. Restor Dent Endod 2015;40(01):91–95
39 Muhammad LN. Guidelines for repeated measures statistical analysis approaches with basic science research considerations. J Clin Invest 2023;133(11):e171058
40 Li H, Johnson T. Wilcoxon's signed-rank statistic: what null hypothesis and why it matters. Pharm Stat 2014;13(05):281–285
41 Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. Korean J Anesthesiol 2016;69(01):8–14
42 Proudfoot JA, Lin T, Wang B, Tu XM. Tests for paired count outcomes. Gen Psychiatr 2018;31(01):e100004
43 Chan Y, Walmslqr RP. Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. Physical Ther 1997;77 (12):1755–1761
44 Sheldon MR, Fillyaw MJ, Thompson WD. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. Physiother Res Int 1996;1(04):221–228
45 McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(02):143–149
46 Kim HY. Statistical notes for clinical researchers: chi-squared test and Fisher's exact test. Restor Dent Endod 2017;42(02):152–155
47 Leon AC. Descriptive and inferential statistics. In: Bellack AS, Hersen M, eds. Comprehensive Clinical Psychology. Oxford: Pergamon; 1998:243–285
48 Goshima S, Kanematsu M, Nishibori H, et al. CT of the pancreas: comparison of anatomic structure depiction, image quality, and radiation exposure between 320-detector volumetric images and 64-detector helical images. Radiology 2011;260(01):139–147
49 Bian Y, Li J, Jiang H, et al. Tumor size on microscopy, CT, and MRI assessments versus pathologic gross specimen analysis of pancreatic neuroendocrine tumors. AJR Am J Roentgenol 2021;217 (01):107–116
50 Millet I, Doyon FC, Hoa D, et al. Characterization of small solid renal lesions: can benign and malignant tumors be differentiated with CT? AJR Am J Roentgenol 2011;197(04):887–896
51 Zhang Y, Zhou Z, Wu C, et al. Population-stratified analysis of bone mineral density distribution in cervical and lumbar vertebrae of Chinese from quantitative computed tomography. Korean J Radiol 2016;17(05):581–589

