1. Bernoulli random variables take (only) the values 1 and 0
    a) True
2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
    b) Central limit theorem

3. Which of the following is incorrect with respect to use of Poisson distribution

    b) Modeling bounded count data

4. Point out the correct statement

    d) All of the mentioned

5. _____ random variables are used to model rates.

    c) Position

6. 10. Usually replacing the standard error by its estimated value does change the CL

    a) false

7. 1. Which of the following testing is concerned with making decisions using data?

    b) hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

    a) 0

9. Which of the following statement is incorrect with respect to outliers
    C) outliers cannot conform to the regression relationship.

10. What do you understand by the term Normal Distribution?

The Normal Distribution is defined by the probability density function for a continuous random variable in a system. Let us say, f(x) is the probability density function and X is the random variable. Hence, it defines a function which is integrated between the range or interval (x to x + dx), giving the probability of random variable X, by considering the values between x and x+dx.

$f(x) \geq 0 \ \forall \ x \in (-\infty, +\infty)$

And $\int_{-\infty}^{+\infty} f(x) = 1$

Normal Distribution Formula

The probability density function of normal or gaussian distribution is given by;

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where,

- x is the variable
- μ is the mean
- σ is the standard deviation

Normal Distribution Curve

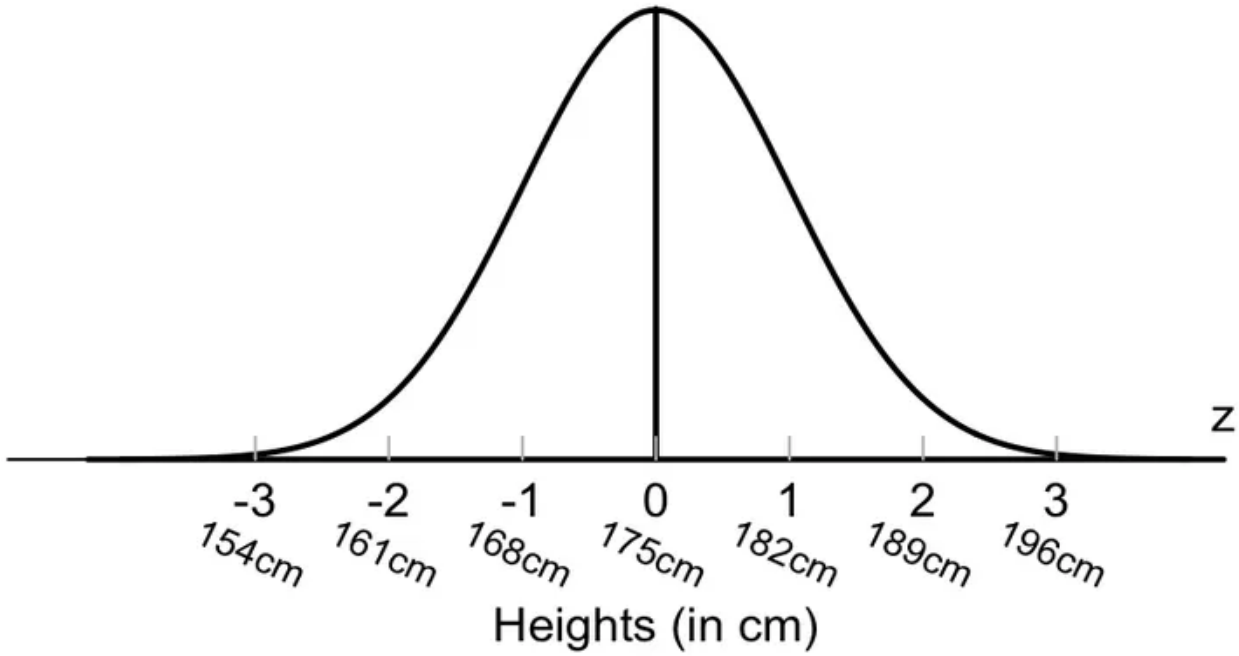
The random variables following the normal distribution are those whose values can find any unknown value in a given range. For example, finding the height of the students in the school. Here, the distribution can consider any value, but it will be bounded in the range say, 0 to 6ft. This limitation is forced physically in our query.

Whereas, the normal distribution doesn't even bother about the range. The range can also extend to –∞ to + ∞ and still we can find a smooth curve. These random variables are called Continuous Variables, and the Normal Distribution then provides here probability of the value lying in a particular range for a given experiment. Also, use the normal distribution calculator to find the probability density function by just providing the mean and standard deviation value.

- Approximately 95% of the data falls within two standard deviations of the mean. (i.e., Between Mean- two Standard Deviation and Mean + two standard deviations)
- Approximately 99.7% of the data fall within three standard deviations of the mean. (i.e., Between Mean- three Standard Deviation and Mean + three standard deviations)

- # Example of a Normal Distribution
- Many naturally-occurring phenomena appear to be normally-distributed. Take, for example, the distribution of the heights of human beings. The average height is found to be roughly 175 cm (5' 9"), counting both males and females.
- As the chart below shows, most people conform to that average. Meanwhile, taller and shorter people exist, but with decreasing

frequency in the population. According to the empirical rule, 99.7% of all people will fall with +/- three standard deviations of the mean, or between 154 cm (5' 0") and 196 cm (6' 5"). Those taller and shorter than this would be quite rare (just 0.15% of the population each).



The normal distribution describes a symmetrical plot of data around its mean value, where the width of the curve is defined by the standard deviation. It is visually depicted as the "bell curve."

11. How do you handle missing data? What imputation techniques do you recommend

Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

Mean imputation

Calculate the mean of the observed values for that variable for all non-missing people.It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.

Substitution

Assume the value from a new person who was not included in the sample.To put it another way, pick a new subject and employ their worth instead.

Hot deck imputation

A value picked at random from a sample member who has comparable values on other variables.To put it another way, select all the sample participants who are comparable on other factors, then choose one of their missing variable values at random.

One benefit is that you are limited to just feasible values. In other words, if age is only allowed to be between 5 and 10 in your research, you will always obtain a value between 5 and 10.Another factor is the random element, which introduces some variation. For exact standard errors, this is crucial.

Cold deck imputation

A value picked deliberately from an individual with similar values on other variables.In most aspects, this is comparable to Hot Deck, but without the random variance. As an example, under the same experimental condition and block, you can always select the third individual.

Regression imputation

The result of regressing the missing variable on other factors to get a predicted value.As a result, instead of utilising the mean, you're relying on the anticipated value, which is influenced by other factors. This keeps the associations between the variables in the imputation model, but not the variability around the anticipated values.

Stochastic regression imputation

The predicted value of a regression plus a random residual value.This has all of the benefits of regression imputation plus the random component's benefits.The majority of multiple imputation is based on stochastic regression imputation.

Interpolation and extrapolation

An estimate based on other observations made by the same person. It generally only works with data that is collected over time.Proceed with caution, though. For a variable like height in children–one that cannot be reduced through time–interpolation would make more sense. Extrapolation entails estimating beyond the data's true range, which necessitates making more assumptions than is necessary.

Single or Multiple Imputation

- Single and multiple imputation are the two forms of imputation. When people say imputation, they usually mean single.
- The term "single" refers to the fact that you only use one of the seven methods to estimate the missing number outlined above.
- It's popular since it's simple to understand and generates a sample with the same number of observations as the complete data set.
- When listwise deletion eliminates a considerable amount of the data set, single imputation appears to be a tempting option. It does, however, have certain restrictions.
- Unless the data is Missing Completely at Random, certain imputation processes, such as means, correlations, and regression coefficients, result in skewed parameter estimations. The bias is frequently worse than with listwise deletion, which is most software's default.
- The level of the bias is determined by a number of factors, including the imputation technique, the missing data mechanism, the fraction of missing data, and the information in the data set.

Furthermore, standard errors are underestimated by all single imputation approaches.Because the imputed observations are estimates, their values have a random error associated with them. However, your programme is unaware of this when you enter that estimate as a data point. As a result, it ignores the additional source of error, resulting in too-small standard errors and p-values.

And, while imputation is straightforward in theory, it is difficult to master in reality. As a result, it isn't perfect, although it may suffice in some circumstances.

As a result of multiple imputation, numerous estimates are generated. In multiple imputation, two of the approaches indicated above–hot deck and stochastic regression–work as the imputation method.

The multiple estimates varied significantly because these two approaches contain a random component. This reintroduces some variance that your program can account for in order to provide reliable standard error estimates for your model.

12. What is A/B testing?

A/B testing definition

A/B testing—also called split testing or bucket testing—compares the performance of two versions of content to see which one appeals more to visitors/viewers. It tests a control (A) version against a variant (B) version to measure which one is most successful based on your key metrics. As a digital marketing practitioner doing either B2B marketing or B2C marketing, your options for conducting A/B tests include:

- Website A/B testing (copy, images, colors designs, calls to action), which splits traffic between two versions—A and B. You monitor visitor actions to identify which version yields the highest number of 1) conversions or 2) visitors who performed the desired action.
- Email marketing A/B testing (subject line, images, calls to action), which splits recipients into two segments to determine which version generates a higher open rate.
- Content selected by editors or content selected by an algorithm based on user behavior to see which one results in more engagement.

Regardless of the focus, A/B testing helps you determine how to provide the best customer experience (CX).
In addition to A/B tests, there are also A/B/N tests, where the "N" stands for "unknown". An A/B/N test is a type with more than two variations.

A/B testing provides the most benefits when it operates continuously. A regular flow of tests can deliver a stream of recommendations on how to fine-tune performance. And continuous testing is possible because the available options for testing are nearly unlimited.

As noted above, A/B testing can be used to evaluate just about any digital marketing asset including:

- emails
- newsletters
- advertisements
- text messages
- website pages
- components on web pages
- mobile apps

A/B testing plays an important role in campaign management since it helps determine what is and isn't working. It shows what your audience is interested in and responds to. A/B testing can help you see which element of your marketing strategy has the biggest impact, which one needs improvement, and which one needs to be dropped altogether.

Benefits of running A/B tests

Website A/B testing provides a great way to quantitatively determine the tactics that work best with visitors to your website. You may simply be validating a hunch, or your hunch could be proven wrong. However, there is still an upside because you won't stick with something that isn't working. You'll attract more visitors who will spend more time on your site and click more links.

By testing widely used website components/sections, you can make determinations that improve not only the test page but other similar pages as well.

13. Is mean imputation of missing data acceptable practice

As a result of reduced varience the model is less accurate and the confidence interval is narrower.

14. What is linear regression in statistics?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

Generate predictions more easily

You can perform linear regression in Microsoft Excel or use statistical software packages that greatly simplify the process of using linear-regression equations, linear-regression models and linear-regression formula. SPSS Statistics can be leveraged in techniques such as simple linear regression and multiple linear regression.

We can perform the linear regression method in a variety of programs and environments, including:

- R linear regression
- MATLAB linear regression
- Sklearn linear regression
- Linear regression Python
- Excel linear regression

Why linear regression is important

Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study.

You’ll find that linear regression is used in everything from biological, behavioral, environmental and social sciences to business. Linear-regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very quickly.

A proven way to scientifically and reliably predict the future

Business and organizational leaders can make better decisions by using linear regression techniques. Organizations collect masses of data, and linear regression helps them use that data to better manage reality — instead of relying on experience and

intuition. You can take large amounts of raw data and transform it into actionable information.

You can also use linear regression to provide better insights by uncovering patterns and relationships that your business colleagues might have previously seen and thought they already understood. For example, performing an analysis of sales and purchase data can help you uncover specific purchasing patterns on particular days or at certain times. Insights gathered from regression analysis can help business leaders anticipate times when their company's products will be in high demand.

15. What are the various branches of statistics

The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.

Descriptive Statistics

Descriptive statistics deals with the presentation and collection of data. This is usually the first part of a statistical analysis. It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.
Different areas of study require different kinds of analysis using descriptive statistics. For example, a physicist studying turbulence in the laboratory needs the average quantities that vary over small intervals of time. The nature of this problem requires that physical quantities be averaged from a host of data collected through the experiment.

Inferential Statistics

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.
Most predictions of the future and generalizations about a population by studying a smaller sample come under the purview of inferential statistics. Most social sciences experiments deal with studying a small sample population that helps determine how the population in general behaves. By designing the right experiment, the researcher is able to draw conclusions relevant to his study.
Both descriptive and inferential statistics go hand in hand and one cannot exist without the other. Good scientific methodology needs to be followed in both these steps of statistical analysis and both these branches of statistics are equally important for a researcher.