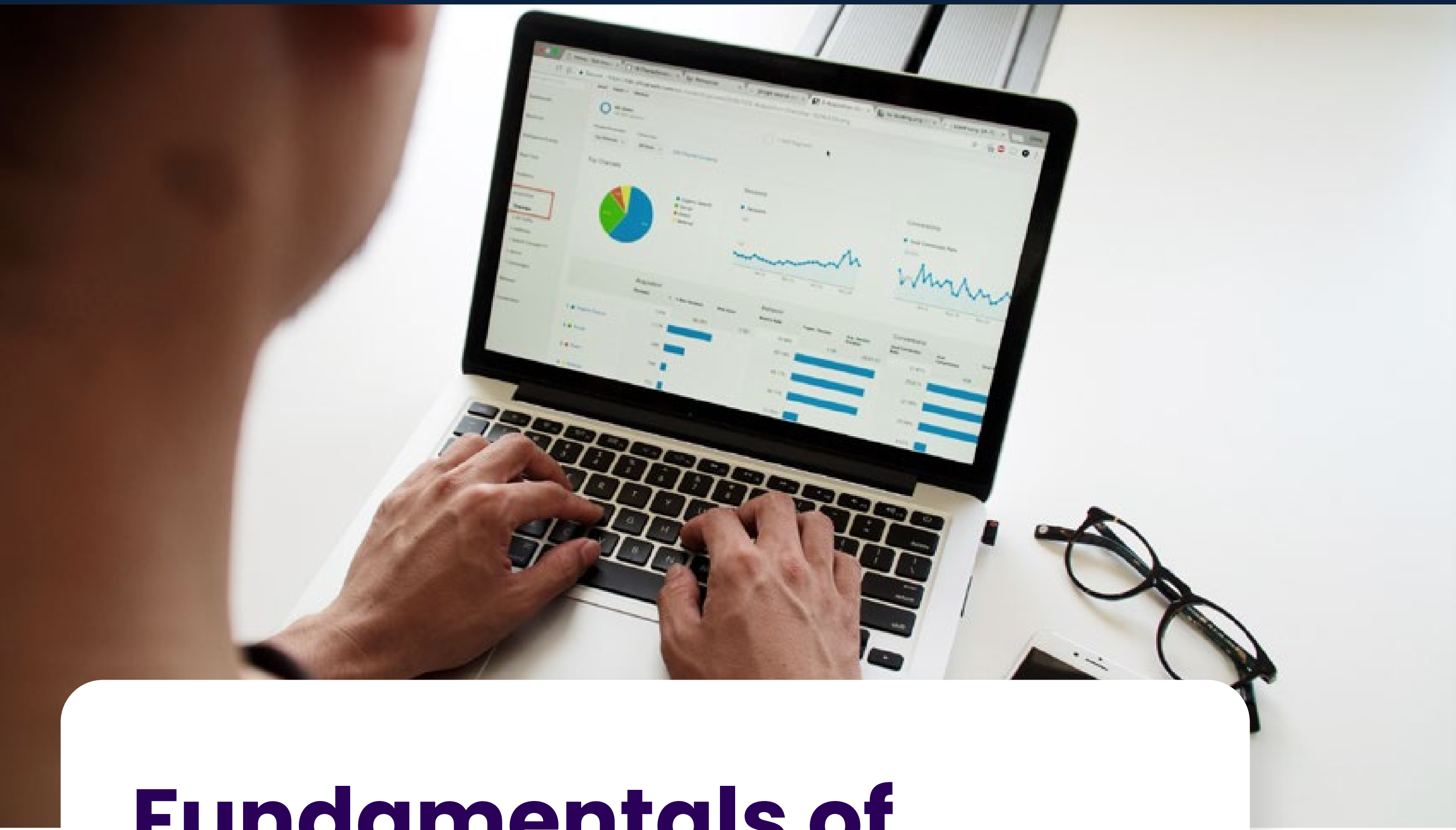
Analyttica TreasureHunt® LEAPS

www.analyttica.com



Fundamentals of Data Analytics

Module 2:

Introduction to Probability Distributions and Statistical Inference

ANALYTTICA TREASUREHUNT
LEARN . APPLY . SOLVE

Table of Contents

Random Variable and Probability Distribution

In the world of uncertainty, output of any event is unknown to us in advance and that makes the decision-making process difficult. It is important to understand the structure of this uncertainty in order to get an idea about risk and reward associated with every action we take. Probability is the quantitative study of uncertainty that helps us make a risk-reward optimized action.

In simple terms, probability refers to chance or likelihood of a particular event taking place. An event is an outcome of an experiment. An experiment is a process that is performed to understand and observe possible outcomes. Set of all outcomes of an experiment is called sample space.

Random Variable

A random variable is a set of possible values, which are the numerical outcomes from a random experiment. For example, consider studying the number of distinct items purchased by a customer visiting a store. We know that the sale value will be any non-negative integer. But we do not know the actual sale value until she finishes her purchase and we look into her basket. Hence, the study is a random experiment. We know what the possible range of outcome could be, but we wouldn't know the actual value of the metric until we measure the same. In this case, the specific sale value for any specific customer is the outcome of the experiment.

Probability Distribution

Though the value of the random variable is not known, the range of possible values are known. From past knowledge, we can assign a probability of the random variable taking a specific value or a range of values.

The distribution of the probability across the range of possible value is known as probability distribution of the random variable.

For example: Let us consider the possible outcome, X, from throwing two unbiased dices together.

Here, X is a random variable that can take any integer value between 2 to 12. Let's try to calculate the probability for each value.

Table 1:

Event	Possible Outcomes	Probability
{X= 2}	{1,1}	1/36
{X=3}	{1,2}, {2,1}	2/36 = 1/18
{X=4}	{1,3}, {3,1}, {2,2}	3/36 = 1/12
{X=5}	{1,4}, {4,1}, {2,3}, {3,2}	4/36= 1/9
{X=6}	{1,5}, {5,1}, {2,4}, {4,2}, {3,3}	5/36
{X=7}	{1,6}, {6,1}, {2,5}, {5,2}, {3,4}, {4,3}	6/36= 1/6
{X=8}	{2,6}, {6,2}, {3,5}, {5,3}, {4,4}	5/36
{X=9}	{3,6}, {6,3}, {4,5}, {5,4}	4/36 = 1/9
{X=10}	{4,6}, {6,4}, {5,5}	3/36 = 1/12
{X=11}	{5,6}, {6,5}	2/36= 1/18
{X=12}	{6,6}	1/36
Total		1

So, total probability, i.e. 1, is distributed among these mutually exclusive and exhaustive events. This is the probability distribution of random variable X.

If the random variable can take only discrete values, e.g. number of accidents, number of computers working, number of transactions, etc., then the variable is known as a discrete random variable, and its corresponding distribution is called a discrete distribution.

If the random variable can take continuous values, e.g. height of an individual, total sales made by a customer, revenue of a company, etc., then the variable is known as a continuous random variable, and the distribution is called a continuous distribution.

Introduction to Discrete Distributions

Probability distributions of discrete random variables are called discrete probability distributions or probability mass functions. These functions assign a probability to each point in the given sample space. There are different types of distributions found in literature, although, we will discuss the two most commonly used distributions, viz. Binomial distribution and Poisson distribution. Prior to that, let's discuss the concept of Bernoulli trials, which defines the random experiment resulting in the corresponding sample space for a Binomial or a Poisson distribution.

Bernoulli Trials

A random experiment is called a Bernoulli trial if it follows the following two conditions:

- The trial has only two possible outcomes – Success or Failure; True or False; Yes or No
- Probability of the outcome of any trial remains fixed over the entire trials
- Trials are statistically independent and random

A common example for a Bernoulli trial is that of flipping a coin. There are only 2 possible outcomes, i.e., heads or tails, with the event of heads being considered a success (or failure) and the event of tails hence being considered a failure (or success). A fair coin has the probability of success 0.5 by definition, since in this case there are exactly two possible outcomes.

Binomial Distribution

Binomial distribution belongs to the family of discrete probability distributions. This is useful to describe the outcome of *n* Bernoulli trials.

Let us assume, *n* independent Bernoulli trials with the probability of a successful trial being *p*. Now, if X is the number of 'success' from the above trials, then *P{X= x}* or *f(x)* is given by:

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad where\ x = 0, 1, 2, ..., n$$

In this case, X is called to follow a Binomial distribution with parameter *n* and *p*.

For binomial distributions, mean of X is *np* and variance is *np(1-p)*

For example: The probability that a driver from a certain demographics will have car accident in a year is 0.3. If 5 drivers from the same demographics are accepted into an insurance program, what is the probability that at most 2 will have an accident?

Solution: To solve this problem, we compute 3 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,
$b(x \leq 2; 5, 0.3) = b(x = 0; 5, 0.3) + b(x = 1; 5, 0.3) + b(x = 2; 5, 0.3)$
$b(x \leq 2; 5, 0.3) = 0.1681 + 0.3601 + 0.3087$
$b(x \leq 2; 5, 0.3) = 0.8369$

So, the probability is 83%.

Poisson Distribution

Poisson distribution can be looked at a special case of Binomial distribution where *n* is very high and *p* is very low. Generally, for rare events like, number of accidents, number of printing mistakes, etc., we assume Poisson distribution.

If μ is the mean occurrence of the event per interval, then the probability of having x occurrences within a given interval is:

$$P(X = x) = \frac{[(e^{-\mu})\,(\mu^{x})]}{x!}$$

In Poisson distribution, the mean and variance are same and equal to μ. Hence, for over dispersion problem, negative binomial distributions are often used.

For Example: The average number of claims that arrive at an insurance firm is 2 claims per day. What is the probability that exactly 3 claims will arrive tomorrow?

Solution: This is a Poisson experiment in which we know the following:

- μ = 2; since 2 claims arrive per day, on average
- x = 3; since we want to find the likelihood that 3 claims will arrive tomorrow
- e = 2.71828; since *e* is a constant equal to approximately 2.71828

We plug these values into the Poisson formula as follows:

$P\ (x; \mu) = (e^{-\mu})\ (\mu^{x}) / x!$
$P\ (3; 2) = (2.71828^{-2})\ (2^{3}) / 3!$
$P\ (3; 2) = (0.13534)\ (8) / 6$
$P\ (3; 2) = 0.180$

So, there is an 18% chance that 3 claims will arrive tomorrow.

Resources for further reading

Binomial Distribution

http://mathworld.wolfram.com/BinomialDistribution.html
http://onlinestatbook.com/2/probability/binomial.html
http://www.stat.yale.edu/Courses/1997-98/101/binom.htm

Poisson Distribution

http://mathworld.wolfram.com/PoissonDistribution.html
https://www.umass.edu/wsp/resources/poisson/
http://www.intmath.com/counting-probability/13-poisson-probability-distribution.php

Introduction to Continuous Distributions

Probability distributions of continuous random variables are called continuous probability distributions or probability density functions. In the case of continuous random variables, we calculate the probability density function instead of the probability mass function as in the case of discrete random variables, since our concentration lies on the probability of an outcome lying within a given interval in the case of continuous random variables. For discrete random variables, our concentration lies on the probability of the outcome being a discrete point in the given space. The integral of the probability density function gives the probability that a random variable falls within some interval.

Normal Distribution or the 'Bell' Curve

The Normal distribution is a very common continuous probability distribution, and is a very important concept in statistics. It is widely used in natural, econometrics and social sciences to represent real-valued random variables, whose distributions are not known.

It is informally known as a bell curve since it resembles a bell in a two-dimensional space.

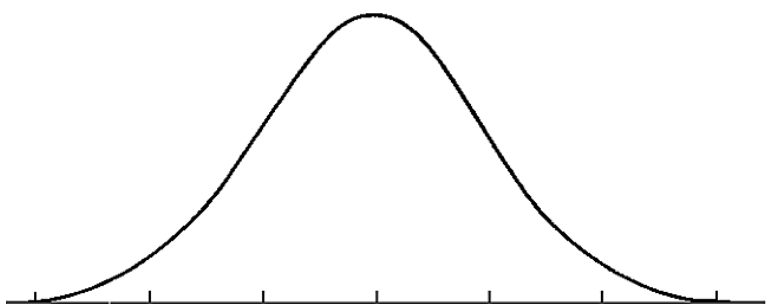


Figure 1: Bell curve

If the probability density function looks as above, we say the variable follows a **Normal distribution**. The probability density function is given as below:

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

We write, X ~ *N (μ, σ²)*

Characteristics of Normal Distribution

The following properties of normal distribution make it the most preferred distribution in any kind of analysis.

- Normal distribution is symmetric around it's mean

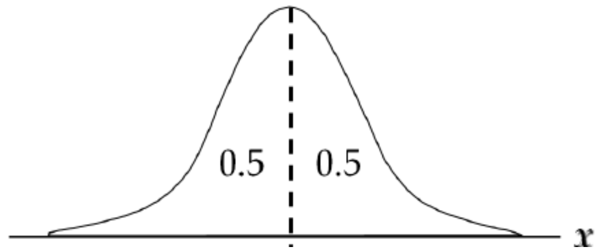


Figure 2: Symmetricity of bell curve

- The variable can take any real number between –∞ to +∞
- Location is determined by μ and shape is determined by σ

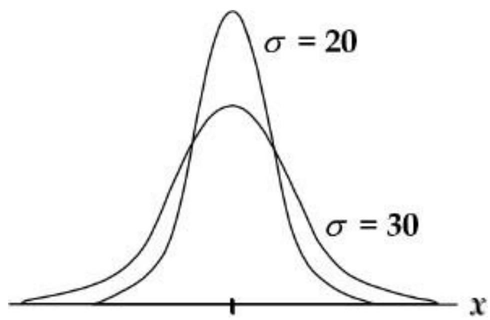


Figure 3: Effect of shape parameter to the bell curve

- Area covered around the mean with multiples of standard deviation is fixed, and this information is used widely across statistical analysis.

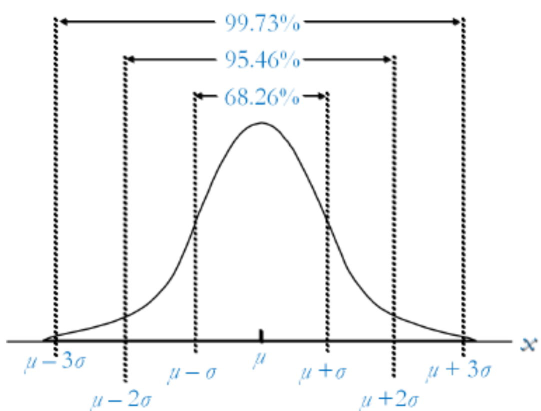


Figure 4: Covered area around the mean

Gamma Distribution

Gamma distribution is relevant to model processes for which the waiting times are relevant. This can be thought equivalent to waiting time between events which are characterized by Poisson distribution.

For example, if we define Y as a random variable that measures number of times a machine fails, then X is expected to follow Poisson distribution. If Z is another random variable that defines time between two consecutive failures of that machine, then Y is expected to follow Gamma distribution.

The probability density function of Gamma distribution is given by:

$$P(x) = \frac{x^{k-1}e^{-x/\theta}}{\Gamma(k)\theta^k}$$

Where,

- The random variable x is a non-negative real number, i.e. $0 \le x < \infty$
- 'e' is the natural number
- 'k' is called the 'shape parameter' and $k > 0$
- 'θ' is called the scale parameter and $\theta > 0$
- 'Γ' is the gamma function. If 'k' is integer, then $\Gamma(k) = (k-1)!$
- $\theta = 1 / \lambda$ is the mean number of events per time unit, where λ is the mean time between events.

For example, if the mean time between trains is 2 hours, then you would use a gamma distribution with θ=1/2=0.5. If we want to find the mean number of trains in 5 hours, it would be 5*(1/2)=2.5. This is also known as scale parameter

For Example: Suppose, you are waiting at a train station and you expect to get a train once every 1/2 hour. Compute the probability that you will have to wait between 2 to 4 hours before 4 trains leave the station. One train every 1/2 hour means we would expect to get θ = 1/0.5 = 2 trains every hour on average. Using θ = 2 and k = 4, we can compute this as follows:

$$P(2 \le X \le 4) = \sum_{x=2}^{4} \frac{x^{4-1}e^{-x/2}}{\Gamma(4)2^4} = 0.12388$$

Resources for further reading

http://mathworld.wolfram.com/Chi-SquaredDistribution.html
http://www.math.uah.edu/stat/special/ChiSquare.html
http://www.r-tutor.com/elementary-statistics/probability-distributions/chi-squared-distribution

Importance of Normal Distribution

In most of the applications of statistical theories, we assume the data to follow normal distributions. This assumption not only makes the application simple, but there is also a strong theoretical justification behind this.

Central Limit Theorem

This is one of the most important theorems in statistics. According to this theorem, the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough.

In other words, if we find the mean of a large number of independent random variables, the mean will follow a normal distribution irrespective of the distribution of the original variables.

In practice, often a sample size of 30 is large enough when the population distribution is roughly bell-shaped. Others recommend a sample size of at least 40. But, if the original population is distinctly not normal (e.g. badly skewed, has multiple peaks, and/or has outliers), researchers like the sample size to be even larger.

Shape Parameters in a Distribution – Skewness and Kurtosis

Till now, we have discussed ‘Measures of Location’ and ‘Measures of Dispersion’ of data. Beyond that, we also often examine the shape of the data or shape of a distribution through two more measures, viz. Skewness and Kurtosis.

In one word, “Skewness” measures the symmetricity of a distribution. For example, let’s look at the distribution of a normal distribution.

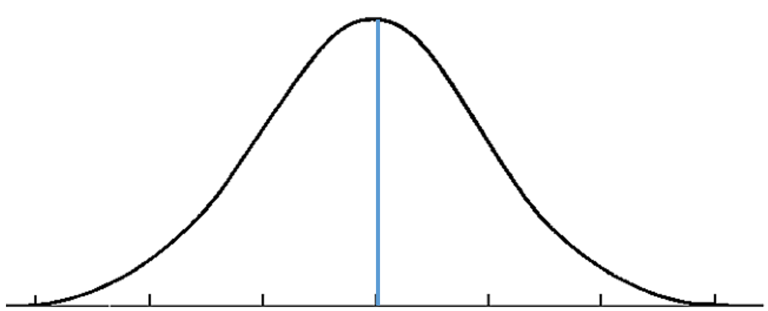


Figure 5: Symmetricity of a Bell Curve

The above graph is symmetric around the central blue line (the left side is a mirror image of the right side). This is called a symmetric distribution. The “skewness” measure for this distribution is zero (0). For any symmetric distribution, Mean = Median = Mode

Now, let's look at the following graph.

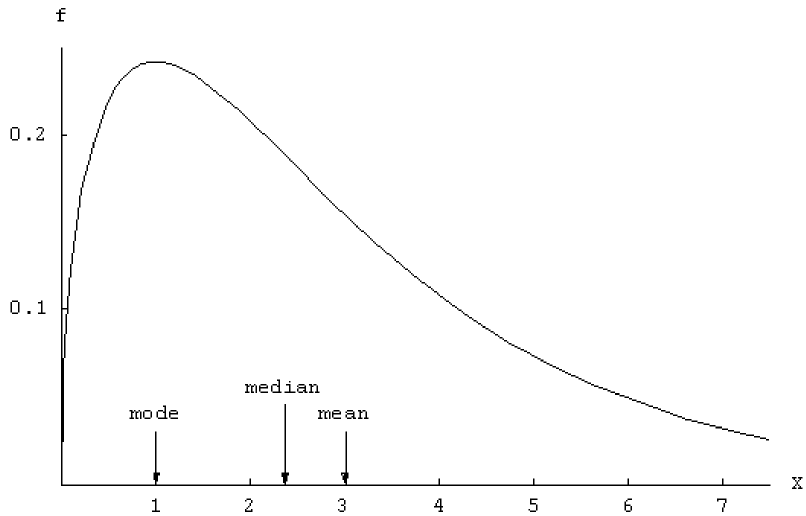


Figure 6: Positively Skewed Distribution

Clearly, it is not symmetric. Here Mean > Median > Mode. We can say that the distribution has a longer tail towards the higher values. Hence, for this kind of asymmetric distributions, skewness measure is positive, and they are called "Positively Skewed Distribution"

Similarly, for negatively skewed distributions Mean < Median < Mode and they have a longer tail towards lower values.

Kurtosis is another measure of shape of a distribution which measures whether the distribution is "flatter" or "thinner" than a normal distribution with same mean and variance.

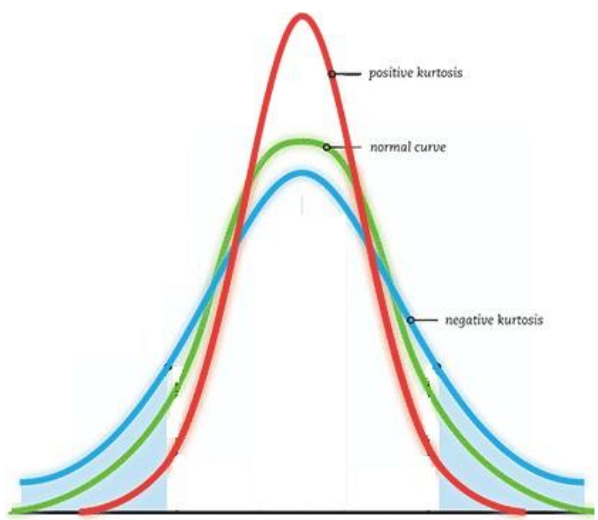


Figure 7: Kurtosis of a Distribution

Non-normality of a Data

Most statistical procedures make an assumption that the variable under consideration (or the error term, to be specific) is Normally distributed. But it is important to check or test if the assumption holds true or not.

We can test if a distribution is normal or not by testing its skewness (it should be close to 0) and kurtosis (close to 3). We can draw a Quantile-Quantile plot. We can also use a statistical test (e.g. Jarque-Bera test).

However, if the underlying distribution is non-normal, we can apply a few transformations to make it normal.

Usually, such transformations include square root/log/inverse transformations. In the next section, we will discuss these transformations individually at greater length.

Square Root Transformation

Here, square root of every value is taken. Special care must be taken to deal with negative numbers as one cannot take the square root of a negative number. Hence in this case, a constant must be added to move the minimum value of the distribution above 0, preferably to 1.00. Also, it is important to remember that numbers above 0.00 and below 1.0 behave differently than numbers 0.00, 1.00 and those larger than 1.00. The square root of 1.00 and 0.00 remain 1.00 and 0.00, respectively, while numbers above 1.00 always become smaller, and numbers between 0.00 and 1.00 become larger (the square root of 4 is 2, but the square root of 0.40 is 0.63). Hence, while applying a square root transformation to a continuous variable that contains values between 0 and 1 as well as above 1, there is a chance that some numbers are being treated differently than others, which may not be desirable.

Log Transformation

This is a class of transformation rather than a single transformation. There are many examples of log-normal variables (i.e. they are normally distributed after log transformation). A logarithm is the power (exponent) a base number must be raised to in order to get the original number. Mathematically, logarithm of numbers less than 0 is undefined, and similar to square root transformations as discussed above, numbers between 0 and 1 are treated differently than those above 1.0. Thus, a distribution to be transformed via this method should be anchored at 1.00.

Inverse Transformation

To take the inverse of a number (x) is to compute 1/x. What this does is essentially make very small numbers (e.g., 0.00001) very large, and very large numbers very small, thus reversing the order of the scores (this is also technically a class of transformations, as inverse square root and inverse of other powers are all discussed in the literature). Therefore, one must be careful to reverse the distribution prior to (or after) applying an inverse transformation. To reverse, one multiplies a variable by -1, and then adds a constant to the distribution to bring the minimum value back above 1.00 (again, as numbers between 0.00 and 1.00 have different effects from this transformation than those at 1.00 and above, the recommendation is to anchor at 1.00).

Other Transformations

Additionally, Arc-Sine transformation can also be applied. It involves taking the arcsine of the square root of a number, with the resulting transformed data reported in radians. Because of the mathematical properties of this transformation, the variable must be transformed to the range −1.00 to 1.00.

All these transformations belong to the family of Box-Cox transformation. This is also known as power transformation.

For example, a square root transformation can be characterized as $x^{1/2}$, inverse transformations can be characterized as x^{-1}, and so forth. Various authors talk about third and fourth roots being useful in various circumstances (e.g. $x^{1/3}$ or $x^{1/4}$).

Box-Cox Transformation can be described as:

$$y_i^\lambda = (y_i^\lambda - 1)/\lambda \; where \; \lambda \neq 0; \; y_i^\lambda = 0$$

How to compute λ by hand:

1. Divide the variable into at least 10 regions or parts
2. Calculate the mean and standard deviation for each region or part
3. Plot log (standard deviation) vs. log (mean) for the set of regions
4. Estimate the slope of the plot, and use the slope (1-b) as the initial estimate of λ

However, most of the statistical packages have in-built or automated ways to compute λ.

Statistical Inference – Estimation and Hypothesis Testing

Now that we are aware of the concepts of sample and population, the next step is to collect sample data and study the population. Hence, our aim is to draw inferences about the population, from the sample data. The method of making conclusions about the population in the purview of the sample data is known as Statistical Inference

There are two broad types of methodologies followed in statistical inference, viz. Estimation and Hypothesis Testing. When we try to conclude about the values of statistics like mean, variance, correlation, median, etc. of a distribution, we call the methodology as "Estimation". On the other hand, "Hypothesis Testing" is about validating some statements about the population.

For example: A plant manager at a cement manufacturing company wants to know the average number of cement bags produced by the plant in an hour. This is a problem of estimation.

Now the manager wants to know if the average number of cement bags produced has increased since a new machine has been installed. This is a problem of hypothesis testing.

Estimation

Estimator and Estimates

Let us consider the above problem, where the plant manager wants to *estimate* the average number of cement bags produced by the plant in an hour. The statistic or metric that is used to measure the value of the population parameter (i.e. mean, median, variance, etc.) is called an estimator.

The value of the estimator from the data is the estimate of the population parameter.

For example: In the above example, the plant manager collected data on 10 hours of production from last 30 working days. They are:

135, 121, 136, 142, 138, 131, 129, 136, 140, 133

Let's assume θ is the population mean. The estimator of θ can be given by:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{10} X_i$$

So, $\bar{X}$ is an estimator of θ.

The value of $\bar{X}$ from the above 10 data points is 134.1. This is the estimate of θ.

Properties of an Estimator

To estimate any population parameter, there could be a number of possible estimators. We choose the best one from those with respect to the following properties of an estimator:

- Unbiasedness – The probability distribution of the estimator is same as that of the population parameter
- Consistency – On increasing the number of data points, the estimate will be closer to the actual data
- Efficiency – The error in the estimated value of parameter is lower than other estimators
- Sufficiency – This means we have used all information in the data about parameter

Type of Estimates

We can give a single value as an estimate to the parameter. For example, the plant manager says the average number of bags produced in an hour is 134.1. This is called point estimate.

Alternatively, the plant manager can say the average number of bags produced in an hour in the plant lies between 132 and 136 with 95% confidence level. This means the estimate lies within a specified interval. This is called an interval estimate.

For an interval estimate, there is always a confidence level attached to it. A 95% confidence interval means there is 95% certainty that the original value will lie in the interval. Higher the value of confidence level wider is the interval.

Hypothesis Testing

In hypothesis testing, we start with a statement or assumption about the population. This forms a hypothesis. Now, we validate this hypothesis based on sample data points. Post that, we either accept the hypothesis to be true or fail to accept the same. This technique is called "Hypothesis Testing".

For example: In a supermarket, the average sale last year was $260. A market researcher wants to know if the average sale has gone up this time or not. So, she randomly selects 25 shoppers from this year and collects their spend records.

The descriptive statistics on the sample looks as below:

Mean = $330.6
Sample Size = 25
Standard Deviation = $154.2
Standard Error of Mean= $30.8

Now, looking at the above, what conclusions can we draw?

Interpreting the Difference

To make any conclusion from the above data, we need to decide if the value $330.6 is significantly different from $260. This difference of ~$70.6 can be due to various reasons.

1. The difference can be due to actual difference in sales behavior. This implies that the average sale has gone up this year compared to previous year
2. The difference can be just due to chance, i.e. the average sales is still the same as last year. But the difference is because we are using a sample instead of population. Another random sample may generate values different from $330.6
3. The third possibility is that there was some design issue in the analysis. For example, the sample was collected from some festive season

By selecting a proper random sample, we try to eliminate the chance of design error. Hence, we need to see if the difference is a true difference or not. A properly designed hypothesis testing enables us to discriminate between two.

Degrees of Freedom

Before getting into further discussions on hypothesis testing, one must understand the concept of "degrees of freedom".

Degrees of freedom, often written as 'df', represents the number of values in the final calculation of a statistic that are free to vary. The number of degrees of freedom is equal to the total number of measurements less the total number of restrictions on the measurements.

Example: Without Constrains

Suppose we have 5 balloons to be given to 5 children. We can only play with four different balloons to decide which one to give to whom. Once 4 balloons are given to 4 different children, we have no choice but to give the remaining to the last child. Hence, the degrees of freedom is 4 (= 5 - 1).

Example: With Constrains

Now, let's add some condition in the above example. Assume there is a pink balloon which we want to give to the youngest child. How many degrees of freedom do we have?

We first give the pink one to the youngest child. Now we are left with 4 balloons and hence 3 degrees of freedom (4 - 1 = 3)

So, degrees of freedom = number of balloons – number of restrictions – 1.

Null and Alternate Hypothesis

In every case of hypothesis testing exercise, there are two hypotheses (or statements) that we evaluate. The hypothesis we test or validate is called "Null Hypothesis". This can be considered as a champion statement, which people believe or expect. "Alternate Hypothesis" is the hypothesis that negates the null hypothesis or challenges the null hypothesis.

Null hypothesis is denoted as H_0 and alternate hypothesis is denoted as H_1. H_1 can completely denote the complement of null hypothesis or just a subset of it.

In any hypothesis testing exercise, the decision maker or the analyst formulates the null hypothesis with a view to get it rejected. This means, she wants the hypothesis to be rejected and hence she tries to validate H_0 only. We do not validate the appropriateness of H_1. Hence, we either say, "There is not enough evidence against null hypothesis. Hence, we accept null hypothesis" Or we say, "There is enough evidence against null hypothesis. Hence we reject null hypothesis". But, rejecting null hypothesis does not mean acceptance of alternate hypothesis.

For example: In the supermarket example, the null hypothesis (H_0) is "there is no change in the average spend" or "Average spend in the current year is \$260". The alternate hypothesis can be simply "there is a change in the average spend", i.e. "Average spend in the current year is NOT \$260". This is a complement set of H_0.

Alternatively, we have reason to believe that the sale cannot go down, then alternate hypothesis becomes "the average spend this year has gone up", i.e. "Average spend in the current year is greater than \$260".

In the supermarket example, the null hypothesis says that there is not significant difference between \$260 and the sample mean, i.e. \$330.6. In other words, it says the difference in sample mean and expected mean is just by chance and no real difference exists.

On the other hand, the alternate hypothesis says the difference exists between previous year sale and current year sale.

Different Types of Errors

Like in any decision-making exercise, errors can occur during hypothesis testing. There are two types of errors that can occur, viz. rejecting H_0 when it is true or accepting H_0 when it is not true. We term these two errors as "Type I Error" and "Type II Error" respectively. We can illustrate this as below:

Decision taken →	Fail to Reject H_0	Reject H_0
Reality: H_0 True	Correct Decision	Type I Error
Reality: H_0 False	Type II Error	Correct Decision

Illustration 1: Two types of errors in hypothesis testing

Ideally, we would like to have a testing method that will minimize both the errors. But unfortunately, this is not possible. As we try to minimize the possibility of occurrence of one error, the same of the other increases.

As a result, we give a special consideration to null hypothesis. We first hold the possibility of occurrence of Type-I error at a very low value and then try to find a test with reasonable low probability of Type II error.

Judicial Analogy

The above idea can be very well analogically compared with judicial system. In any judicial system, we give special consideration to the fact that "an accused is innocent". Until we are reasonably sure, we do not punish the accused to ensure "no innocent should be punished, even if we have to free some guilty"

Similarly, until the evidence against null hypothesis is sufficiently strong, we do not reject it. Hence, we try to reduce the chance to reject null hypothesis when it is true, i.e. Type I error.

Significance Level and p-Value

The level where we want to fix the probability of Type I error is known as the level of significance or simply significance level of the test. We denote this by the Greek letter α (alpha).

P(Type I Error) = Level of Significance = α

Generally, we fix the value of α at a very low value like 0.05 or 0.01

Now, based on the data we calculate something called 'p-Value', which indicates given the dataset, probability that the null hypothesis is true. If the p-value is lower than significance level, then we reject the null hypothesis. Else we accept the null hypothesis.

For example: In the supermarket data, we do a t-test (to be discussed in the next module). The p-value for the test comes out to be 0.012. This means that there is only 1.2% likelihood that the null hypothesis (i.e. there is no change is average sale this year) is true. In other words, only 12 out of 1,000 cases we can expect the sample mean to be $330.6 or more when population mean is $260.

If we consider α= 0.05, then we conclude that there is enough evidence against null hypothesis. But if we consider α= 0.01, then we conclude that there is not enough evidence against null hypothesis.

Write to us at

support@analyttica.com

USA Address

Analyttica Datalab Inc.

1007 N. Orange St, Floor-4, Wilmington, Delaware - 19801

Tel: +1 917 300 3289/3325

India Address

Analyttica Datalab Pvt. Ltd.

702, Brigade IRV Centre,2nd Main Rd, Nallurhalli,

Whitefield, Bengaluru - 560066.

Tel : +91 80 4650 7300