Normal Distribution, “p” Value and Confidence Intervals

NJ Gogtay, SP Deshpande, UM Thatte

When data is collected, in order to make sense of it, the data needs to be organised in a manner which shows the various values and the frequencies at which these values have occurred, that is the “pattern that the values form after they are organised” and this is called a distribution. A distribution guides how raw data can be converted into meaningful information and also influences the choice of the appropriate statistical tests so that correct conclusions may be drawn.

The “Normal Distribution” is probably the most important and most widely used distribution in statistics. It is also called the “bell curve” or the “Gaussian” distribution after the German mathematician Karl Friedrich Gauss (1777–1855). It is useful to note that the word “Normal” does not indicate that it is “normal” (as against abnormal) and statisticians use a capital N to emphasise this. Although many biological phenomena are Normally distributed, in some specialties in medicine, Normal distributions are rare e.g. oncology.

A Normal distribution has several properties that make it useful for inferential statistics. These include the following:

1. Every Normal distribution is characterized by its mean and standard deviation.
2. The area under the Normal curve is 1.0 (100%) and is divisible for the purpose of analysis as in point 3.
3. The area of one SD on either side represents 68% of the population, 2SD on either side 95% of the population and 3SD 99% of the population.
4. The distribution is dense at the centre and less dense at its tails.
5. It describes continuous data.

The data is spread evenly and equally around the central value (which is the “mean”) with 50% of data falling on either side of the mean. For example, if there is a sample of values of HbA1C levels of 1000 people, which is “Normally distributed”, it means that there are 500 values below and 500 above the mean (Figure 1). Importantly, the mean, median and mode are very close to each other.

If the data is Normally distributed, then it should always be described as mean ± SD. Conversely, if the data is not Normally distributed it should NEVER be described as mean ± SD. Rather, median and range/interquartile range are used to describe this type of data. Further, parametric tests can be used to analyse Normally distributed data while if the data is not Normally distributed, then non-parametric tests of significance should be employed to find a statistical difference.

Tests to Assess Normality of Distribution

Prior to expressing data as mean/median and to decide what tests can be applied to analyse data for statistical significance, it is necessary to identify whether the data is Normally distributed or not (for the reasons described above). Several tests are available to assess Normality and include the Kolmogorov-Smirnov (K-S) test, Lilliefors corrected K-S test,

![Fig. 1: Normal distribution](image-url)
When Data is not Normally Distributed

There are a large number of distributions that describe data as shown in Table 1 that are not “Normal”. Just like the data which they describe, distributions can be classified as continuous or discrete. These distributions are useful for performing specific analysis of data and some of these will be discussed in the forthcoming articles (e.g. Binomial Distribution for binary categorical data in Survival Analysis, Chi Square Distribution in Chi Square test, t distribution in the t-test, and so on).

Although the sub-type of distribution influences the test of significance chosen, conventionally, data distributions that are not Normal, are not actually tested for the type of distribution prior to applying a statistical test.

When data is not Normally distributed we can either transform the data (for example, by taking logarithms) and apply parametric tests or use non-parametric tests which do not require the data to be normally distributed for analysis.

**“p” Value**

The first computations of p-values were calculated by Pierre-Simon Laplace as far back as the late 18th century. At around the same time, Laplace considered the statistics of almost half a million births. The statistics showed an excess of boys compared to girls. He concluded by calculation of a p-value that the excess was a real, but unexplained, effect.

The p-value was first formally introduced by Karl Pearson and noted as capital P. The use of the p-value in statistics was popularized by Sir Ronald Fisher who proposed the level $p = 0.05$, or a 1 in 20 chance of being exceeded by chance, as a limit for statistical significance. Today, the success or failure of our intervention hinges on this ubiquitously quoted “p” value.

In order to appreciate this concept, let us take an example of two drugs used for hypertension. You believe that Drug A brings about a greater fall in blood pressure than does Drug B. We would study a group of patients treated with drug A and a comparable group treated with Drug B and find a reduction that is greater in the group treated with drug A. This is encouraging but could it be merely a chance finding? We examine the question by calculating a “p” value: the probability of getting at least a 2 mmHg difference in BP reduction. When you plan the clinical study, you start by making a statement that there is no difference between the two drugs in the BP reduction.

This is called the “null” hypothesis ($H_0$). When the study is completed and you find that Drug A indeed brings about a greater fall in BP, you “reject” the null hypothesis or “accept” the “alternative hypothesis, ($H_1$)”, which is that there is a difference between the two drugs in the BP reduction they produce. We always apply a statistical test to find out whether there is a difference between Drug A and B. This test (and in the next article we will see which tests are useful and where) throws up a “p” value. Also called the “probability” value, this number tells us whether the observed difference is a “true” difference or is occurring simply by chance. Conventionally, a “p” value less than 5% is considered to be “significant”. This means that in our example above, if we get a value of $p<0.05$ (5%) it means that the probability that Drug A brings about a greater fall in BP than drug B is $>95$% and that this effect was purely due to chance alone is <5%. It is important to remember that this value does NOT give the magnitude of the difference i.e. the extent of BP reduction may be only marginally different between the two drugs and yet we may get a significant p value if the effect is consistent across the sample. Also if the p value is $<0.001$, it only means that there is a stronger evidence of the effect being NOT due to chance (Figure 2). However, it is at the discretion of the investigator to set what p value will be considered significant (but it has to be less than 5%). This should be decided at the time of planning the protocol and is influenced by strength of evidence desired for rejecting $H_0$, the number of groups being compared and whether the study will have interim analyses.

When publishing, “p” values must be presented as actual values (like $p=0.012$ or $p=0.002$) rather than merely stating that $p<0.05$ or $p<0.01$. A “significant” p value should be interpreted in the context of the type of study and other

---

**Table 1: Types of distributions other than normal distributions**

<table>
<thead>
<tr>
<th>Continuous</th>
<th>Discete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>Binomial</td>
</tr>
<tr>
<td>t distribution</td>
<td>Poisson</td>
</tr>
<tr>
<td>Cauchy</td>
<td>Gamma</td>
</tr>
<tr>
<td>F distribution</td>
<td>Power normal</td>
</tr>
<tr>
<td>Chi square</td>
<td>Power log normal</td>
</tr>
<tr>
<td>Exponential</td>
<td>Tukey lambda</td>
</tr>
<tr>
<td>Weibull</td>
<td>Extreme value type 1</td>
</tr>
<tr>
<td>Log normal</td>
<td>Beta distribution</td>
</tr>
<tr>
<td>Birnbaum Saunders</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Double exponential</td>
<td></td>
</tr>
<tr>
<td>Power normal</td>
<td></td>
</tr>
<tr>
<td>Power log normal</td>
<td></td>
</tr>
</tbody>
</table>

Although the sub-type of distribution influences the test of significance chosen, conventionally, data distributions that are not Normal, are not actually tested for the type of distribution prior to applying a statistical test.

When data is not Normally distributed we can either transform the data (for example, by taking logarithms) and apply parametric tests or use non-parametric tests which do not require the data to be normally distributed for analysis.
available evidence, as well as its clinical relevance.

**Confidence Interval**

For practical purposes, we study “samples” and extrapolate results to the population from where the sample is drawn. The purpose of statistical inference is to estimate population parameters using observed sample data without having to actually study the whole population. A 95% confidence interval (95% CI) is a population parameter expressed as a range that contains the true population parameter 95% of the times. 2.5% of the values will be below and 2.5% of the times above this range. Here again, in medicine, it is conventional to use a (95% CI) rather than 90, 97 or 99 or any other but this is only a convention. If your calculated population parameter must be as close to certain as possible, you would calculate the (99% CI) while if you knew that there would be heterogeneity and a smaller sample size, you could measure a (90% CI) as is done in Bioequivalence studies.

Let us take the example of the study of Drug A vs. Drug B for reduction in BP. If the mean difference between the reduction in BP produced by Drug A and B is 5mmHg with a 95% CI of [2, 8], then this is interpreted to mean that the difference in the reduction in BP seen in the population will fall between 2 and 8mm Hg 95% of the times. It is useful to note that this range is expressed in square brackets with a comma between the upper and lower bounds. Confidence intervals can be computed for various parameters, and not just the mean.

The next chapter will describe the common statistical tests used in health research.

**Reference**