Statistical Problems in the Determination of Normal Ranges from Laboratory Data*

Hae Hiang Song Ph. D.

Department of Preventive Medicine & Biostatistics, Catholic Medical College

═국문초록═

임상 검사 결과로 부터 정상 범위 추정에 대한 통계학적 연구

가톨릭의과대학 예방의학 · 통계학 교실

송   혜   향

질병의 진단이나 예후를 판정함에 도움이 되고자 의학분야에서 정상 범위를 정의해 왔으며, 이 정의에 도입된 전통적인 통계적 방법으로는 정규 확률지를 사용하여 정상범위를 추정함이었다. 이 방법은 정상군과 질병군의 검사 성적 분포가 각각 정규 분포한다는 가정에서 시작되며, 이 가정이 보편적일 수 없음을 여러 연구(Elveback et al, 1970; Feinstein, 1977)에서 밝히고 있다. 또한 통계적 방법론중 일상 사용되고 있는 평균치로부터 두배의 표준 편차 거리로 나타 내어지는 구간을 정상 범위로 추정함은 비대칭 분포의 경우 합당치 못한 경우가 생기게 된다. 대칭과 비대칭의 분포에 합당하게 사용될 수 있는 방법은 백분율에 근거한 방법으로 특정한 의학적인 요구가 없는 한 균등한 하단과 상단의 백분율 구간을 정함이 상례였다고 보겠다. 그러나 높은 칼슘의 수치를 나타내 보이는 환자에게 주의가 필요함과 같이 특수한 의학적인 측구에 의해 하단과 상단의 불균등한 백분율 지점을 정상 범위로 선정할 수도 있으나 백분율의 표준 오차에 근거해서 정상 구간이 추정되어야 하겠다.

A false premise

Graphical aids are often of particular use to statisticians. A special graph called normal probability paper determine whether the distribution of a given sample is approximately normal. This is done by plotting the cumulative distribution of a sample upon the probability paper and then noting how closely this curve approximates a straight line. If the curve is approximately a straight line, the distribution is approximately normal. If it deviates conside-rably from a straightline, the distribution is not normal.

In modern medicine normal probability paper is introduced in determining normal range by Hoffmann (1963), Waid (1964) and Neumann (1968). The basic premise for their use of probability paper is that the observed distribution of routine laboratory tests is a mixture of two normal (or gaussian) distributions, one representing the healthy persons and the other, the sick, as is shown in Fig 1. The composite distribution of two normals is asymmetric and from a plot on normal probability paper the line corresponding to the healthy group is deduced by capitalizing that most of the specimens sent to laboratories for analysis are normal. On a
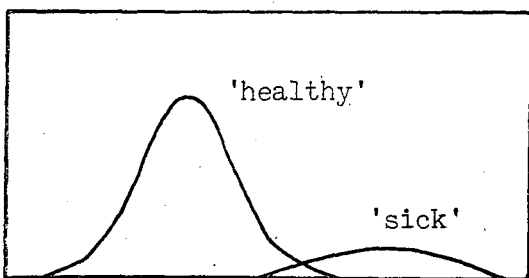
Fig. 1. Two theoretical gaussian distributions, one representing the healthy and the other the sick individuals.

straight line 2.5% and 97.5% points are established graphically. This eye-fitting is so unreliable that it may provide diverse normal ranges depending on who evaluates. It is reported by Elveback and others (1970) that this graphical method of determining normal range is in fairly wide use.

One way of testing the reasonableness of the above premise is to actually test several hundred normal persons and check the shape of distributions. In a study of Elveback and others (1970) values of serum calcium, inorganic phosphorus, magnesium, total proteins, albumin and urea for 576 healthy persons are plotted. The distributions in healthy persons are not gaussian for the majority of tests, and the departures from normality may be major.

Asymmetric distributions arise naturally in biological contexts. For example, the distribution of the number of cones on a fir trees in a given area of forest, or the distribution of the number of lepidoptera of the same species present and observed in a particular location, are both characterized by high skewness. Over the years it became well established that the normal distribution is inadequate to represent certain set of biological data. In medicine, however, it is not firmly convinced that discordant values are in no way anomalous, in general. The graphical method of establishing normal range by normal probability paper is to be

dismissed in medicine for the future, since except in a few instances there may be no reason to believe the underlying distribution is normal.

Normal range

It is certain that normal range has a comforting role in the first phase of the diagnostic process: the finding of a value outside the range, and sometimes even within but near border-line, usually requires closer examination of the patient. Eventually practicing physicians might diagnose patients by overall clinical and morphologic observations. Let us pause for a moment to compare the medical meaning of 'normal' and the statistical statement of 'normal'. There are many articles written in the past opposing the use of normal range. It is considered that the misunderstanding on normal range is due to the lack of distinction between these two terms.

A single value is necessary but insufficient in determining that an individual is normal and thus medical diagnosis of normality usually need to incorporate diverse aspects of human beings, possibly requiring a careful examination. This should not, however, defer adopting quantification and other scientific procedures of analyzing values of a test. If we contemplate the statistical normal range by eliminating any direct association with medical phenomena, it is not correct either. Pure statistical statements shoud be understood in the meaning of average, common, and conventional; in other words, it should be understood in the meaning of probability. If the particular value is excluded from the univariate statistical normal range, it is because that particular value implies the 'higher possibility' of the hazard of some ailment. In accepting this it also implies that there is a small chance of unduly diagnosing healthy in-

dividuals, but with grossly abnormal test values, as diseased.

Normal range defined by $\bar{x} \pm 2s$

In an article of The Lancet (1967) the normal values are defined as those within two standard deviations from the mean of a population. Krupp and Chatton (1984) adopt this definition and present a list of normal ranges of chemical constituents of blood and body fluids. Is the definition discussed above appropriate as the definition of normal values? What is the condition with the distribution of actual test values that the normal range encompasses approximately 95% of the distribution? If the distribution is gaussian, statistical theory explains that it is symmetric and that 95% is contained within two standard deviations from the mean of the
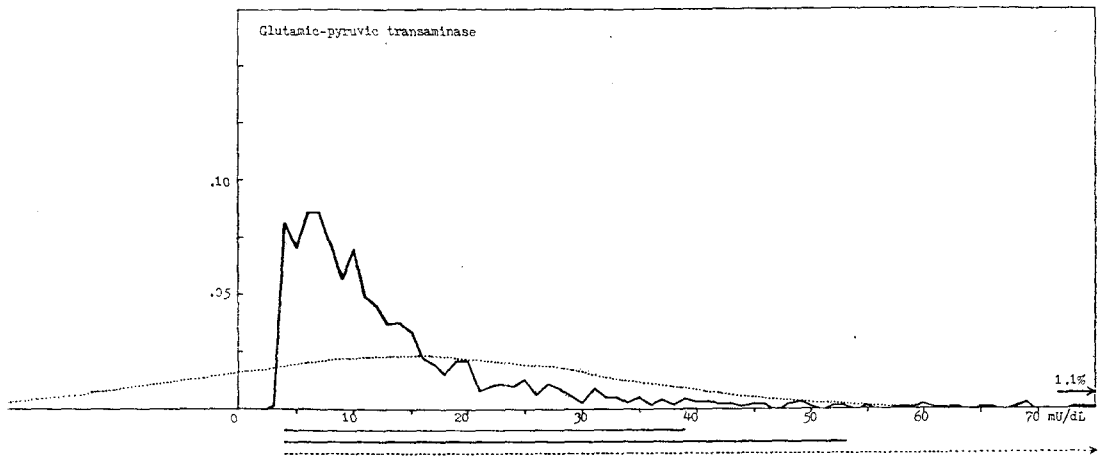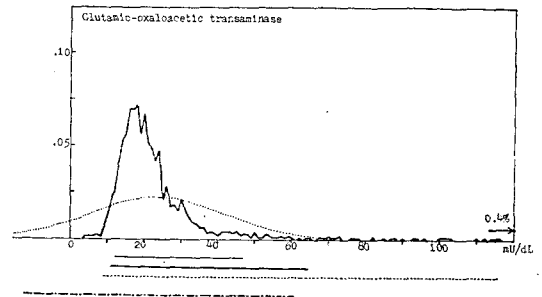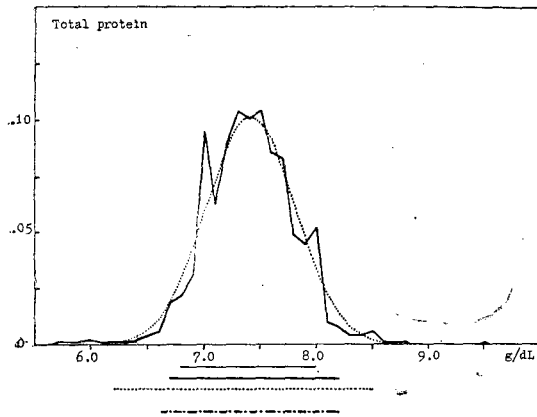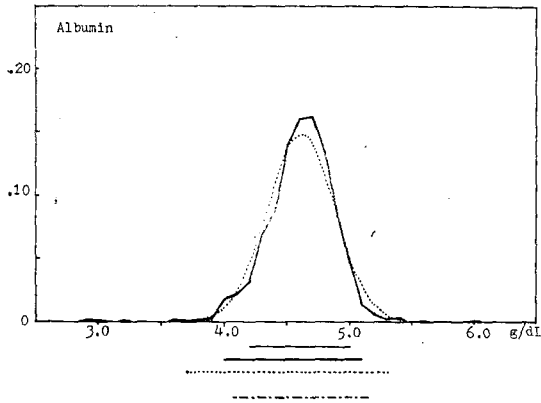

Fig. 2. Distributions of the hepar functions of 1,847 persons examined during January-April, 1983 at the Computerized Health Screening Center of Kangnam St. Mary's Hospital. (The central ranges by the percentile method, 5th to 95th, $2\frac{1}{2}$th to $97\frac{1}{2}$th and $\frac{1}{2}$th to $99\frac{1}{2}$th, and the range defined by $\bar{x} \pm 2s$ is drawn in its order.)

distribution. Thus, altogether 5% of abnormally low or high values are excluded by the defined range of $\bar{x} \pm 2s$, where $\bar{x}$ is the sample mean and s the sample standard deviation.

To investigate a gaussian assumption for the distribution of actual test values, results of hepar functions are obtained for 1,847 persons examined during January-April, 1983 at the Computerized Health Screening Center of Kangnam St. Mary's Hospital (Samples were taken from individual after fasting, with no drinking water allowed, from 10 PM of previous night). The distributions of serum albumin, total protein, glutamic-oxaloacetic and glutamic-pyruvic transaminase enzyme test values are chosen and presented in Fig 2. For transaminase enzyme tests, deviations from the gaussian curve are fairly obvious, but the distributions of total protein and albumin appear to be of gaussian shape. When in fact the assumption is tested for albumin and total protein, the results show a statistically significant difference from a gaussian curve ($p < 0.01$). There are many ways in which a distribution of test values can differ from a gaussian curve. The distribution of albumin is sharply peaked (leptokurtic) compared to normal and that of total protein deviates from normal as a result of few humps.

The use of $\bar{x} \pm 2s$ as the definition of normal range, strictly speaking, is justified only in the case where the distribution is close to a gaussian or at least a symmetric distribution. In other cases it may be used to make statements in probability but stating that roughly 95% of the distribution is contained within the range $\bar{x} \pm 2s$ rely in fact on a gaussian assumption. For each test gaussian distribution is fitted using the sample mean and the standard deviation (Fig 2). For total protein the normal range turns out to be (6.627, 8.197) and the proportion of sample values lower than 6.627 and higher than 8.197 is, respectively, 1.8% and 2.6%, totalling

4.4% For albumin the normal range is (4.069, 5.145) and the proportion of values lower than 4.069 and higher than 5.145 is, respectively, 2.8% and 1.4%, totalling 4.2%. Although the distributions are not exactly gaussian, use of range $\bar{x} \pm 2s$ in dissecting the central 95% of the distribution is appropriate since they are symmetric.

The absurdity of the normal range defined by $\bar{x} \pm 2s$ for transaminase enzyme tests is obvious. The range $\bar{x} \pm 2s$ corresponds to (-12.535, 59.685) for the test of glutamic-oxaloacetic and to (-19.889, 49.679) for glutamic-pyruvic. For both tests the range extends to, biologically impossible, negative values, and hence the sample proportion of values less than 0 is null for both tests and the proportion of sample values higher than the upper limit is respectively 2.8% for glutamic-oxaloacetic and 2.8% for glutamic-pyruvic. In fact the proportion of values less than 0 is considerably higher for both tests under the fitted normal distributions. It is thus inevitable to resort to some other definitions of normal range for asymmetric distributions.

The reason that the range $\bar{x} \pm 2s$ plays an important role in many fields at present time is due to the superior mathematical tractability and certain optimal sampling properties of the mean and the standard deviation. Another reason is due to the commonly mistaken notion that the distributions of most test values are gaussian. It should be remembered that for the case of asymmetric distributions the defined normal range is quite misleading sometimes.

Normal range defined by percentiles

No matter in which classes the distributions might be fall into, the principal characteristics of the distribution is the measure of location and the measure of dispersion. There are three measures of location in common use: the mean,

the median and the mode. And, as the measures of dispersion, there are the standard deviations, the mean deviations, the range, the interdeclie range and the interquartile range. The ranges are the differences of certain lower and upper values and they are in general less known as measures of dispersion. Not only the range is of interest: the lower and upper percentage points themselves help locating the distribution. The range itself determining the degree of scatter and the whereabouts of these lower and upper points explaining the location of the distribution, these measures evidently give some idea of the distribution.

The apparent conflict discussed earlier between a normal range defined by $\bar{x} \pm 2s$ and biological reality has made researchers turn to a percentile method. This percentile method is generally recognized as the best method currently available for the determination of the normal range by many (Harwood et al (1978), Elveback et al (1970) and Feinstein (1977)). Some arbitrary limit of two percentiles, commonly 5th to 95th percentile, is chosen as the normal range.

Percentiles of $\frac{1}{2}$th to $99\frac{1}{2}$th, $2\frac{1}{2}$th to $97\frac{1}{2}$th and 5th to 95th are obtained and drawn in Fig 2 for the distribution of hepar functions. The range defined by $\bar{x} \pm 2s$ is also added. For the symmetric distributions of total protein and albumin the range $\bar{x} \pm 2s$ is shown to be close to the range of $2\frac{1}{2}$ th to $97\frac{1}{2}$th percentiles and, as the percentages within the range decreases, the width of the range tends to narrow fast. For the asymmetric distributions of transaminase enzyme tests, though the range defined by $\bar{x} \pm 2s$ provides an unreasonable limits, the percentile ranges are appropriate except that the range of $\frac{1}{2}$th to $99\frac{1}{2}$th percentiles may seem to be too wide and with that limit a substantial number of individuals would be classified as normal. It is doubtful how Krupp and Chatton (1984) has determined, based on the very definition of

$\bar{x} \pm 2s$, the normal ranges of various tests done for the blood and body fluids, since it is unwarranted that the distributions of the tests presented by them were all gaussian.

Certainly percentile method has an advantage that it is obtained irrelevant to the shape of a distribution. Hence, ranges defined by the percentiles are always safe. Nevertheless, this method has very little use currently. Percentiles suffer from the disadvantage of being relatively difficult to handle mathematically compared to the mean and the standard deviation, since they are not expressed as algebraic functions of test values.

Standard error of percentiles

For the percentile method there seem to be not an enough and convincing explanation why the statistical strategy devides symmetrically on either side of the distribution. Why choose an equal break of $2\frac{1}{2}\%$ and $2\frac{1}{2}\%$, for example, for the case of 95% range? This question is to be investigated based on the standard error of the percentiles. If the lower and upper percentiles are taken to be equal, we say that the ranges are central. In the contrary case the ranges will be called non-central. It should be observed that centrality in this sense implies the equal distribution of percentiles at lower and upper extremes.

If p_1 $(=1-q_1)$th percentile of a sample of size n falls below a value x_1, the sampling variance of x_1 is

var $x_1 = p_1 q_1 / n f^2_1$,

where f_1 is the ordinate of the population distribution at x_1. (Refer Kendall & Stuart, 1967). Thus, the variance of the precentile point will be unknown unless the population distribution is specified. The variance of range R is the variance of the difference of two percentiles at x_1 and x_2 and is given by

$$\text{var } R = \text{var} x_1 + \text{var} x_2 - 2\text{cov}(x_1, x_2)$$

$$= \frac{1}{n} \left\{ \frac{p_1 q_1}{f_1^2} + \frac{p_2 q_2}{f_2^2} - \frac{2p_2 q_1}{f_1 f_2} \right\}$$

where f_1, f_2 are the ordinate of the population distribution, f_2 relating to the upper percentile.

Suppose the population distribution is symmetric and furthermore normal with variance σ^2. Then $f_1 = f_2$ and the standard error of the central range of $2\frac{1}{2}$th to $97\frac{1}{2}$th percentiles is $4.037\sigma/\sqrt{n}$, since the ordinate at both percentile points is $f_1 = .053991/\sigma$. What about the non-central range? The standard error of the non-central range of 1st to 4th percentiles, for instance, is obtained as $4.140\sigma/\sqrt{n}$, since $f_1 = .02635$ and $f_2 = .08650$. It turns out that the standard error of the central and non-central ranges do not differ much for the case of gaussian population, resulting somewhat smaller values for the central range.

If the population distribution is normal, cen-tral range of $2\frac{1}{2}$th to $97\frac{1}{2}$th percentiles obtained from a sample of test values would approximately equal to the range obtained by $\bar{x} \pm 2s$. However, standard errors of the ranges determined by two different methods are not the same, since standard error is related to the way in which parameters are estimated. At the present time, the standard errors of these two methods are being investigated by the present author.

It is of interest to consider how the standard errors of the range differ for the asymmetric population distribution. There are many distributions classified as asymmetric. Among them chi-square distribution is unimodal and has an unlimited range in one direction. The standard errors of the range are studied when the population distribution is specified as chi-square. Chi-square distribution with four and six degrees of freedom are shown in Fig 3 and the standard

Table 1. Standard errors of the central and non-central percentile ranges when the population distribution is specified as chi-square with 6 degrees of freedom.

percentiles of the range	Standard error $\times \sqrt{n}$			Standard errors of the range		
	upper percen-tile point	lower percen-tile point	range	$n=100$	$n=200$	$n=300$
$2\frac{1}{2}\% \sim 97\frac{1}{2}\%$	3.020	16.430	16.629	1.290	1.084	.980
$2\% \sim 97\%$	3.087	15.108	15.344	1.239	1.042	.941
$1\% \sim 96\%$	3.249	13.228	13.556	1.164	.979	.885
$5\% \sim 95\%$	2.944	11.921	12.127	1.101	.926	.837
$4\% \sim 96\%$	2.975	10.970	11.236	1.060	.892	.806
$3\% \sim 97\%$	3.000	10.220	10.511	1.025	.862	.779
$2\frac{1}{2}\% \sim 97\frac{1}{2}\%$	3.020	9.915	10.298	1.015	.853	.771

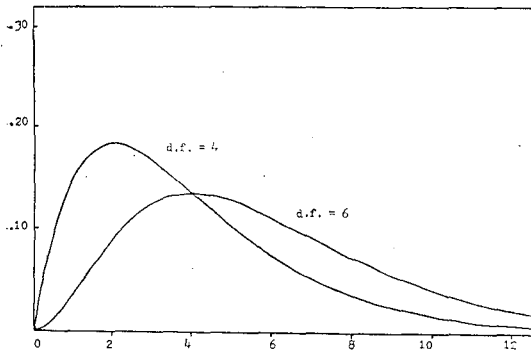| percentiles of the range	Standard errors of the range						
	$n=400$	$n=500$	$n=1,000$	$n=2,000$	$n=3,000$	$n=4,000$	$n=5,000$
$2\frac{1}{2}\% \sim 97\frac{1}{2}\%$	.912	.863	.725	.610	.551	.513	.485
$2\% \sim 97\%$	.876	.828	.696	.586	.529	.493	.466
$1\% \sim 96\%$	.823	.778	.655	.550	.497	.463	.438
$5\% \sim 95\%$	.778	.736	.620	.521	.470	.438	.415
$4\% \sim 96\%$	.750	.709	.596	.501	.453	.422	.399
$3\% \sim 97\%$	.725	.686	.576	.485	.438	.407	.386
$2\frac{1}{2}\% \sim 97\frac{1}{2}\%$	.718	.679	.571	.480	.434	.404	.382

Fig. 3. Chi-square distributions with four and six degrees of freedom.

·errors of the range for different percentiles are shown in Table 1 for the population distribution of chi-square with six degrees of freedom. The standard errors of the upper percentile points are getting larger as the point is further away from the origin and that of the range itself depends largely on the magnitude of the standard error corresponding to the upper percentile point. Chi-square distribution extends to positive infinity and the lack of precision for upper percentile point is to be expected. As the sample size increases, the standard error decreases. The 95% central range with the sample size of 400 has the same magnitude of standard error with the 90% central range of sample size of a little over 200. And, the 95% central range with the sample size of 4,000 has the same magnitude of standard error with the 90% central range of sample size of a little over 2,000.

Different reliability is expected for the range of percentiles if the population distribution is of some other families. And the result informs us that unless the population distribution is specified most reliable range is unknown in general. In the absence of other considerations it might be convenient to employ central ranges, but circumstances sometimes arise in which non-central ranges are more serviceable. If clinical interest centers largely on elevated values

for certain tests, for instance, calcium, then we must clearly err, on the basis of experiences to date, on the safe side by imposing higher percentiles to the upper extremes. In the contrary case in which the accuracy of the lower limit of the range is more desired than that of the upper limit, higher percentiles are imposed to the lower extremes.

Summary

Normal range has use mainly in the first phase of the diagnostic process, that is, to screen or to raise ideas about possible pathology. The traditional method of determining it is based on the probability paper or on the mean plus or minus two standard deviations. These methods are often turned out to be vague and impractical. The percentile method is adequate and flexible, though. The appropriate limit of lower and upper points should be chosen by considering medical aspects above all things and also the reliability of the range determined by the standard error.

The results of normal range are interesting, strictly speaking, only for the hospital concerned. Differences exist between the normal ranges reported by various sources (Bezemer et al, 1983). It would be best to establish the normal range based on a population comparable to a group of individuals to whom the normal range is to serve as a norm.

REFERENCES

1. Bezemer, P.D., Netelenbos, J.C., Mulder, C., Theune, J.A., Stamhuis, I.H. and Straub, J.P.: *Determining Reference ('Normal') Limits in Medicine: An Application, Statistics in Medicine* 2: 191-198, 1983.
2. Elveback, L.R., Guillier, C.L. and Keating, F.R. Jr.: *Health, Normality, and the Ghost of Gauss,* J.A.M.A., 211(1):69-75, 1970.

3. Feinstein, A.R.: *The Derangements of the 'Range of Normal', Clinical Biostatistics, The C.V. Mosby*, 1977.

4. Harwood, S.J. and Cole, G.W.: *Reference Values Based on Hospital Admission Laboratory Data, J.A.M.A.*, **240**(3):270-274, 1978.

5. Hoffmann, R.G.: *Statistics in the Practice of Medicine, J.A.M.A.*, **185**(11):864-873, 1963.

6. Kendall, M.G. and Stuart, A.: *The Advanced Theory of Statistics, Volume 1, Hafner Publ., Co.*, 1967.

7. Krupp, M.A. and Chatton, M.J.: *Current Medical Diagnosis & Treatment*, p. 1085-1103, 1984.

8. Leading Articles: *Interpretation of Laboratory Tests, The Lancet*, **1**:1091-1092, 1967.

9. Neumann, G.J.: *Determination of Normal Ranges from Routine Laboratory Data, Clin. Chem*, **14**: 979-988, 1968.

10. Waid, M.: *Quoted by Sparapani, A. and Berry, R.E.: The Range of Normal Values in the Quality Control of Clinical Chemistry, Am. J. Clin. Path.*, **42**:133-136, 1964.