Statistical tolerance intervals and regions

Thomas Mathew
Department of Mathematics and Statistics
University of Maryland Baltimore County (UMBC)
Baltimore, Maryland 21250

mathew@umbc.edu

IABS 10th Annual Statistics Workshop

November 12, 2024

Part 1: Motivating examples, definitions, different types of tolerance intervals, tolerance intervals under a univariate normal distribution, normal-based tolerance intervals for the lognormal and gamma distributions.

Part 2: Non-parametric tolerance limits, tolerance limits under linear regression models.

Part 3: Tolerance regions and reference regions under the multivariate normal distribution, tolerance rectangles, applications in laboratory medicine.

Not covered in the lectures: Detailed theoretical derivations, non-normal continuous distributions, random effects models, Bayesian approaches, sample size determination, discrete distributions.

Basic reference:

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications and Computation*. John Wiley.

R package:

Young, D. S. (2010). `tolerance`: An R package for estimating tolerance intervals. *Journal of Statistical Software*, 36, 1-39.

Young, D. S. (2020). `tolerance`: Statistical tolerance intervals and regions. R Package, Version 2.0.0.

Examples discussed will include environmental, engineering and clinical applications.

Part 1: Motivating examples, definitions, different types of tolerance intervals, tolerance intervals under a univariate normal distribution, normal-based tolerance intervals for the lognormal and gamma distributions.

Confidence intervals and prediction intervals have only one percentage associated with them, namely, the confidence level, often denoted by $1 - \alpha$.

A standard choice of $1 - \alpha$ is 95% or 0.95.

A tolerance interval has two percentages associated with it. One of them is called the content, to be denoted by p. The second percentage is a confidence level, to be denoted by $1 - \alpha$.

Common values of the content p and the confidence level $1 - \alpha$ are 90%, 95% and 99% (or 0.90, 0.95 and 0.99).

For example, we will talk about a tolerance interval having content $p = 0.90$ and confidence level $1 - \alpha = 0.95$.

The content and confidence level can be equal.

An example

Air lead levels ($\mu g/m^3$) collected by the National Institute of Occupational Safety and Health (NIOSH) at a laboratory, for health hazard evaluation. The air lead levels were collected from 15 different areas within the facility.

It was noted that a normal distribution fitted the log-transformed lead levels quite well (that is, the data follows a lognormal distribution).

Air lead levels

200	120	15	7	8	6	48	61
380	80	29	1000	350	1400	110	

Log-transformed air lead levels:

5.298, 4.787, 2.708, 1.946, 2.079, 1.792, 3.871, 4.111, 5.940, 4.382, 3.367, 6.908, 5.858, 7.244, 4.700

The occupational exposure limit (OEL) for lead exposure set by the Occupational Safety and Health Administration (OSHA) is 50 $\mu g/m^3$, which is 3.912 on the log-scale.

Can we use the log-transformed data to conclude that 90% or more of the air lead levels in the facility are below 3.912?

Note: We want to conclude if 90% or more of the population values are below 3.912. The 90% is not associated with a random sample.

The 90% is the content p. We have $p = 0.90$. The content refers to a percentage for the population.

Note: It is not enough that the average of the log-transformed air lead levels in the facility is below the limit of 3.912.

X: log-transformed air lead levels.

μ and σ^2: population mean and variance for X. $X \sim N(\mu, \sigma^2)$.

Let $\overline{X}$ and S denote the sample mean and sample standard deviation based on a sample of size n.

A 95% confidence interval for μ: $\overline{X} \pm t_{n-1;0.975}\frac{S}{\sqrt{n}}$

A 95% upper confidence bound for μ: $\overline{X} + t_{n-1;0.95}\frac{S}{\sqrt{n}}$

A 95% lower confidence bound for μ: $\overline{X} - t_{n-1;0.95}\frac{S}{\sqrt{n}}$

To predict the air lead level at a particular area within the laboratory.

A 95% upper prediction limit for the log-transformed lead level:

$$\overline{X} + t_{n\text{-}1;0.95}\, S\sqrt{1+\frac{1}{n}}\,.$$

A lower prediction limit, or a two-sided prediction interval can be similarly computed.

The confidence limits and prediction limits just have only one percentage associated with them, namely the confidence level, equal to 0.95. The 95% captures the sampling variability.

To conclude whether or not at least 90% of air lead levels in the population are below the threshold of 3.912.

The confidence interval and prediction interval cannot answer this question.

What is required is a **tolerance interval**; more specifically, an **upper tolerance limit**.

The upper tolerance limit will be computed using the sample mean $\overline{X}$ and the sample SD S.

Recall: A 95% upper confidence bound for μ: $\overline{X} + t_{n-1;0.95} \frac{S}{\sqrt{n}}$

We shall take the upper tolerance limit to be of the form $\overline{X} + kS$

The quantity k is referred to as the tolerance factor for the upper tolerance limit.

Since $\overline{X} + kS$ is a random quantity, the percentage of the population air lead levels below $\overline{X} + kS$ is a random variable. Thus we cannot guarantee that at least 90% of air lead levels in the population is below $\overline{X} + kS$.

However, we can determine k so that at least 90% of air lead levels in the population is below $\overline{X} + kS$ with a high probability, say 95%.

Two associated percentages: 90% and 95%.

The two percentages, 90% and 95%, can be interpreted as follows:

(i) the content 90% (or more) represents the percentage of the population that is required to be below the upper tolerance limit of $\bar{X} + kS$.

(ii) the confidence level 95% represents the probability for meeting the requirement in (i). Reflects sampling variability.

As already noted, the content is denoted by p: $p = 0.90$

The confidence level is denoted by $1 - \alpha$: $1 - \alpha = 0.95$

Solution to the tolerance factor k:

Let z_p denote the pth percentile of the standard normal distribution. For $p = 0.90$, $z_p = 1.28$

The value of k depends on the $100(1-\alpha)$th percentile of a non-central t distribution with df = $n-1$ and non-centrality parameter $= z_p\sqrt{n}$

Let $t_{n-1;1-\alpha}(z_p\sqrt{n})$ denote this percentile.

Then $k = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n})$

For $n = 15$, $p = 0.90$ and $1 - \alpha = 0.95$, $t_{n-1;1-\alpha}(z_p\sqrt{n}) = t_{14;0.95}(4.957) = 8.002$

$$k = \frac{1}{\sqrt{n}}\, t_{n-1;1-\alpha}(z_p\sqrt{n}) \;= \frac{1}{\sqrt{15}} \times 8.002 = 2.066$$

For a normal distribution with mean μ and SD σ, the pth percentile is given by $\mu + z_p\sigma$, where z_p is the pth percentile of the standard normal distribution.

The upper tolerance limit $\overline{X} + kS$ is also an upper confidence limit for $\mu + z_p\sigma$, having confidence level $1 - \alpha$.

For $n = 15$, $p = 0.90$ and $1 - \alpha = 0.95$, $\overline{X} + (2.006 \times S)$ is a 95% upper confidence limit for $\mu + 1.28\sigma$.

Thus the upper tolerance limit can also be interpreted as an upper confidence limit.

Example (continued)

Log-transformed air lead levels:

5.298, 4.787, 2.708, 1.946, 2.079, 1.792, 3.871, 4.111, 5.940, 4.382, 3.367, 6.908, 5.858, 7.244, 4.700

Sample mean = $\bar{X}$ = 4.333, Sample SD = S = 1.739

To decide whether 90% or more of the air lead levels in the population are below 3.912

Let us compute an upper tolerance limit having content $p = 0.90$ and confidence level $1 - \alpha = 0.95$

The upper tolerance limit is given by
$\bar{X} + kS = 4.333 + 2.066 \times 1.739 = 7.926$.

Interpretation?

Note that the upper tolerance limit is not below the threshold of 3.912.

Thus we cannot conclude that 90% or more of the air lead levels in the population are below 3.912.

Lower tolerance limit

Suppose we want to use the sample data to compute a limit so that at least 90% of the population is more than the limit with a high probability, say 0.95.

Such a limit is referred to as a lower tolerance limit, and is given by $\bar{X} - kS$, for the same k:

$$k = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n})$$

For the air lead level example with $n = 15$, $\overline{X} = 4.333$, $S = 1.739$, $k = 2.066$ when $p = 0.90$ and $1 - \alpha = 0.95$

Thus the lower tolerance limit is
$\overline{X} - k \times S = 4.333 - 2.066 \times 1.739 = 0.740$

Interpretation?

For a normal distribution with mean μ and SD σ, the $(1-p)$th percentile is given by $\mu - z_p\sigma$, where z_p is the pth percentile of the standard normal distribution.

The lower tolerance limit $\bar{X} - kS$ is also a lower confidence limit for $\mu - z_p\sigma$, having confidence level $1 - \alpha$.

Thus the lower tolerance limit can also be interpreted as a lower confidence limit.

A clinical example: two-sided tolerance intervals

Serum creatinine levels are used to assess how well the kidneys are working.

The Mayo Clinic web site reports the following ranges for the serum creatinine levels for healthy men and for women.

Adult men: 65.4 to 119.3 micromoles/L

Adult women: 52.2 to 91.9 micromoles/L

The ranges maybe different for different populations, and may also dependent on age (in addition to gender).

Laboratory medicine

From a certain healthy population, measurements of the serum creatinine levels were made for a sample of 284 healthy individuals.

Sample mean $\bar{X} = 85$ and sample SD $S = 14.34$.

We want to determine a range that will include the serum creatinine levels for 90% or more of the adults in the healthy population.

Taken from the book
Harris, E. K. and Boyd, J. C. (1995). *Statistical Bases of Reference Values in Laboratory Medicine*. Marcel Dekker.

The range will be determined using the serum creatinine levels for the sample of 284 healthy individuals.

Thus we are looking for a two-sided tolerance interval having content $p = 0.90$.

The interval is required to contain at least 90% of the serum creatinine levels for the population from which the sample was selected.

Since a sample was used to compute the tolerance interval, a confidence level $1 - \alpha$ will be associated with it. We may choose $1 - \alpha = 0.95$.

The two-sided tolerance interval is taken to be $\bar{X} \pm kS$. That is, the interval $(\bar{X} - kS, \ \bar{X} + kS)$.

The factor k is to be determined subject to the condition that the interval will include the serum creatinine levels of at least 90% of the healthy adults in the population, with a confidence level of 0.95.

There is no closed form expression available for the factor k. However, tables are available. Approximations are also available.

The two-sided tolerance interval does not have a confidence interval interpretation.

Some values of the two-sided tolerance factor k for confidence level $1 - \alpha = 0.95$ and content $p = 0.90, 0.95$, and 0.99

	Content p		
n	0.90	0.95	0.99
11	2.275	2.815	3.852
12	2.210	2.736	3.747
13	2.155	2.671	3.659
14	2.109	2.614	3.585
15	2.068	2.566	3.520
16	2.033	2.524	3.464
17	2.002	2.486	3.414
18	1.974	2.453	3.370
19	1.949	2.423	3.331
20	1.926	2.396	3.295

An accurate approximation is available for the two-sided tolerance factor k:

$$k \simeq \left[\frac{(n-1) \times \chi^2_{1;p}(1/n)}{\chi^2_{n-1;\alpha}} \right]^{1/2}$$

$\chi^2_{n-1;\alpha}$: the (100α)th percentile of a chisquare distribution with df = $n-1$.

$\chi^2_{1;p}(1/n)$: the 100pth percentile of a noncentral chisquare distribution with df = 1 and noncentrality parameter $1/n$.

Approximate and exact two-sided tolerance factors: a = approximate; b = exact

	$1-\alpha = 0.90$				$1-\alpha = 0.95$			
	p				p			
	0.90		0.99		0.95		0.99	
n	a	b	a	b	a	b	a	b
3	5.85	5.79	8.97	8.82	9.92	9.79	12.9	12.7
4	4.17	4.16	6.44	6.37	6.37	6.34	8.30	8.22
5	3.49	3.50	5.42	5.39	5.08	5.08	6.63	6.60
6	3.13	3.14	4.87	4.85	4.41	4.42	5.78	5.76
7	2.90	2.91	4.52	4.50	4.01	4.02	5.25	5.24
8	2.74	2.75	4.27	4.27	3.73	3.75	4.89	4.89
9	2.63	2.64	4.10	4.09	3.53	3.55	4.63	4.63
10	2.55	2.55	3.96	3.96	3.38	3.39	4.43	4.44

For the creatinine example, $n = 284$, $\bar{X} = 85$ and $S = 14.34$. For $p = 0.90$ and $1-\alpha = 0.95$,

$$k \simeq \left[\frac{(n-1) \times \chi^2_{1;p}(1/n)}{\chi^2_{n-1;\alpha}} \right]^{1/2}$$

$$= \left[\frac{283 \times 2.715}{245.037} \right]^{1/2}$$

$$= 1.771$$

The two-sided tolerance interval for the creatinine levels in the population is $\bar{X} \pm kS$. That is $85 \pm (1.771 \times 14.34)$.

Simplifies to the interval (59.6, 110.4).

Interpretation?

Creatinine ranges from the Mayo Clinic web site:

Adult men: 65.4 to 119.3 micromoles/L

Adult women: 52.2 to 91.9 micromoles/L

For $p = 0.95$ and $1 - \alpha = 0.95$,

$$k = \left[\frac{283 \times 3.855}{245.037} \right]^{1/2} = 2.11$$

The two-sided tolerance interval for the creatinine levels in the population is $\bar{X} \pm kS$. That is $85 \pm (2.11 \times 14.34)$.

Simplifies to the interval (54.7, 115.3). Interpretation?

Creatinine ranges from the Mayo Clinic web site:

Adult men: 65.4 to 119.3 micromoles/L

Adult women: 52.2 to 91.9 micromoles/L

The concept of a reference interval

Extensively used in laboratory medicine

The creatinine example: A reference interval for creatinine is defined as the interval that contains 95% of the central part of the distribution for a healthy population, termed the reference population.

The interval is taken to be the interval from the 2.5th to the 97.5th percentiles of the reference population, so that the interval contains 95% of the central part of the distribution.

The computation of the reference interval is based on a random sample from the reference population, and is subject to sampling variability, quantified using a coverage probability.

With a high probability, the estimated reference interval should cover 95% of the central part of the population distribution.

In other words, the estimated reference interval should include the 2.5th and the 97.5th percentiles of the reference population, with a high probability.

The reference interval so computed is referred to as a central tolerance interval.

It is not enough to simply estimate the 2.5th to the 97.5th percentiles from the sample. Why not?

If the reference population follows a normal distribution with mean μ and SD σ, then the 2.5th and 97.5th percentiles are $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$, respectively.

Thus a central tolerance interval is an interval computed using a random sample so that the interval contains the interval $(\mu - 1.96\sigma, \mu + 1.96\sigma)$ with a high probability, say 0.95.

What is the difference between a tolerance interval and a central tolerance interval?

The stronger requirement makes the central tolerance interval wider.

In a published User's Guide on reference intervals, Horne and Pesce (2005, p. 1) note that "The reference interval is the most widely used medical decision-making tool. It is central to the determination of whether or not an individual is healthy".

Horn, P. S. and Pesce, A. J. (2005). *Reference Intervals: A User's Guide*, American Association for Clinical Chemistry Press, DC.

National Committee for Clinical Laboratory Standards (2010). *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline, Third Edition*.

The central tolerance interval is assumed to be of the form $\overline{X} \pm kS$.

The central tolerance factor k is to be computed so that the interval $(\overline{X} - kS, \overline{X} + kS)$ will contain the interval $(\mu - 1.96\sigma, \mu + 1.96\sigma)$ with probability 0.95.

Tables are available for computing the central tolerance factor.

Thus there are two types of two-sided tolerance intervals: with and without the "central" requirement.

Comparison of the tolerance factor and central tolerance factor:
$p = 0.95$, $1 - \alpha = 0.95$

n	Tolerance factor	Central tolerance factor
11	3.273	3.568
12	3.175	3.456
13	3.093	3.363
14	3.024	3.284
15	2.965	3.216
16	2.913	3.157
17	2.868	3.104
18	2.828	3.058
19	2.793	3.016
20	2.760	2.978

Note: The central tolerance factor is always larger than the tolerance factor for a two-sided tolerance interval. Thus a central tolerance interval will always be wider than the two-sided tolerance interval.

For the creatinine example, $n = 284$, $\bar{X} = 85$ and $S = 14.34$.

Take $p = 0.95$, $1 - \alpha = 0.95$

Tolerance factor = 2.11, Tolerance interval: (54.7, 115.3)

Central tolerance factor = 2.174

Central tolerance interval (reference interval): (53.8, 116.2)

Difference between the tolerance interval and central tolerance interval in terms of interpretation?

Creatinine ranges from the Mayo Clinic web site:

Adult men: 65.4 to 119.3 micromoles/L

Adult women: 52.2 to 91.9 micromoles/L

Normal-based tolerance intervals for the lognormal distribution

Let's recall the first example on air lead levels:

Air lead levels (sample from a lognormal distribution):

Air lead levels							
200	120	15	7	8	6	48	61
380	80	29	1000	350	1400	110	

Log-transformed air lead levels (sample from a normal distribution):

5.298, 4.787, 2.708, 1.946, 2.079, 1.792, 3.871, 4.111, 5.940, 4.382, 3.367, 6.908, 5.858, 7.244, 4.700

Sample mean = $\bar{X}$ = 4.333, Sample SD = S = 1.739

Upper tolerance limit having content = p = 0.90 and confidence level = $1 - \alpha$ = 0.95: $\bar{X} + kS = 4.333 + 2.066 \times 1.739 = 7.926$.

This is the upper tolerance limit for the log-transformed air lead levels in the population.

Hence the upper tolerance limit for the air lead levels in the population is exp(7.926) = 2768.

If the sample is from a lognormal distribution, we can take the natural logarithm of the data, compute the tolerance limits using the formulas for the normal distribution, and then take the exponential in order to get the tolerance limits for the original population.

Normal-based tolerance intervals for the gamma distribution

An example: Data on alkalinity concentrations in ground water (mg/L). Obtained from the site of a waste disposal landfill. A sample of 27 observations.

58	82	42	28	118	96	49	54	42	51
66	89	40	51	54	55	59	42	39	40
60	63	59	70	32	52	79			

The gamma distribution fits the data well.

We want to compute an upper tolerance limit for the alkalinity concentrations with content 0.90 and confidence level 0.95.

Taken from the book
Gibbons, R. D., Bhaumik, D. K. and Aryal, S. (2009). *Statistical Methods for Groundwater Monitoring*, 2nd Edition, Wiley

Let $X \sim \texttt{Gamma}(a, b)$, the gamma distribution with shape parameter a and scale parameter b.

The gamma density is given by

$$f(x) = \frac{b^a}{\Gamma(a)} x^{b-1} e^{-bx} \quad .$$

If X follows a gamma distribution, then $X^{1/3}$ has an approximate normal distribution.

Known as the Wilson-Hilferty approximation.

Also known as the cube root normal approximation.

$X_1, X_2, \ldots, X_n$: sample from a gamma distribution.

Let $Y_i = X_i^{1/3}, \;\; i = 1, \ldots, n$.

Assume the Y_i's form a random sample from a normal distribution.

Compute tolerance limits based on the Y_is using the formulas applicable to the normal distribution.

Then take the third power of the tolerance limits so computed in order to get the tolerance limits for the original population.

Data on alkalinity concentrations:

58	82	42	28	118	96	49	54	42	51
66	89	40	51	54	55	59	42	39	40
60	63	59	70	32	52	79			

For $Y_i = X_i^{1/3}$, $i = 1, \ldots, 27$, sample mean = $\overline{Y} = 3.8274$, SD = $S_y = 0.4298$.

For $p = 0.90$, $1 - \alpha = 0.95$, the tolerance factor k under the normal distribution is

$$k = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n}) = \frac{1}{\sqrt{27}} t_{26;0.95}(6.6511) = 1.81$$

Upper tolerance limit:

$$U = \overline{Y} + k \times S_y = 3.8274 + 1.81 \times 0.4298 = 4.6051.$$

The upper tolerance limit for the alkalinity concentrations is $U^3 = 97.66$.

A comment on the accuracy of the cube root normal approximation:

If the shape parameter is very small, the gamma distribution is highly skewed, and the cube root normal approximation is not very accurate.

However, the upper tolerance limit is still reasonably accurate, but the lower tolerance limit is not so.

Tolerance interval calculations based on the cube root normal approximation, and their accuracy, are investigated in:

Krishnamoorthy, K., Mathew, T. and Mukherjee, S. (2009). Normal-based methods for a gamma distribution: prediction and tolerance intervals and stress-strength reliability. *Technometrics*, **50**, 69–78.

Gamma distributions with small shape parameters do come up in many applications; for example, in modeling rain rates. Accurate likelihood based procedures are available for computing tolerance limits in such cases.

Bebu, I. and Mathew, T. (2016). Environmental data analysis based on the gamma distribution: compliance assessment using tolerance limits and exceedance fractions. *Environmetrics*, 27, 370-376.

Summary of Part 1

Introduced the concept of tolerance intervals

Depends on two percentages: content and confidence level

Discussed different types of tolerance intervals: upper tolerance limit, lower tolerance limit, two-sided tolerance interval, central tolerance interval.

Discussed their computation for the normal, lognormal and gamma distributions.

Part 2: Non-parametric tolerance limits, tolerance limits under linear regression models.

Non-parametric tolerance limits

The tolerance intervals discussed so far assumed that the sample is from a normal, lognormal, or gamma distributions.

Parametric assumptions, leading to parametric tolerance intervals

Parametric tolerance intervals can be computed if the data come from some parametric distribution.

If parametric distributional assumptions do not hold, we have to use a nonparametric approach, which does not rely on any assumption concerning the underlying probability distribution.

However, the nonparametric approach is applicable only when the data come from a continuous distribution.

An application: A problem of interest in the evaluation of nuclear power plant safety.

A major focus in the evaluation of nuclear power plant safety is the *loss of coolant accident (LOCA)*.

From the U.S. Nuclear Regulatory Commission site (www.nrc.gov):

LOCA: "Those postulated accidents that result in a loss of reactor coolant at a rate in excess of the capability of the reactor makeup system from breaks in the reactor coolant pressure boundary, up to and including a break equivalent in size to the double-ended rupture of the largest pipe of the reactor coolant system".

In the nuclear reactor, nuclear fuel pellets are stacked into metallic tubes, and then sealed.

The sealed tubes are called fuel rods, which are used to build up the core of a reactor.

The outer protective layer (very often made of steel) between the nuclear fuel and the coolant is referred to as *cladding*.

If the peak cladding temperature (PCT) remained below 2200°F, the cladding would not become embrittled, resulting in fragmentation.

It is important to verify if most of the PCT distribution is below 2200°F.

One recommendation: obtain PCT data from the outputs of computer codes used to simulate nuclear accidents

Compute a non-parametric upper tolerance limit having content and confidence level both equal to 0.95.

If such an upper tolerance limit is below 2200°F, conclude that most of the PCT distribution is below 2200°F.

A data set provided by the company Framatome, Inc.

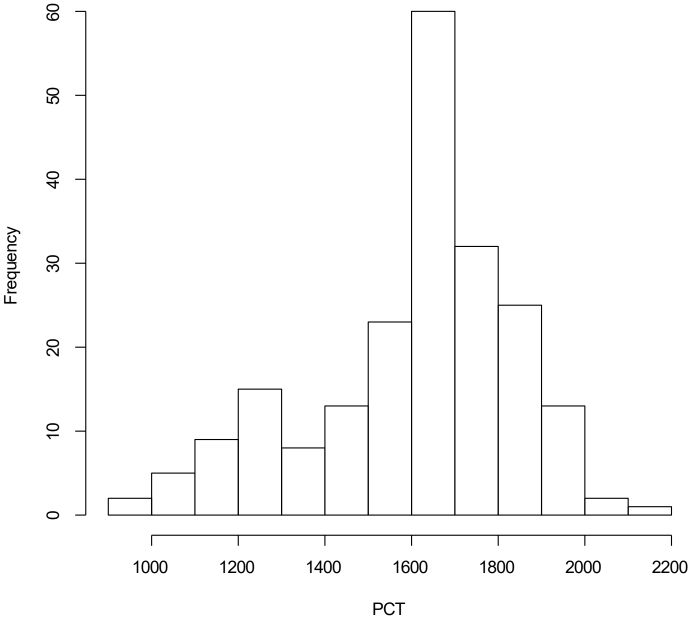
French company, world's largest nuclear company, specializing in power generation.

Over 35 locations in the United States, headquartered in Charlotte.

A few observations (PCT values) in the sample:

1615.2, 1799.8, 1742.4, 1624.1, 1682, 1213.5, 1613.1, 1626.7, 1633.4, 1647.8, 1098.8, 1516.4, 1678.4, 1621.6, 1619, 1069, 1508.2, 1050.2, 1738.2, 974.8, 1658.2, 1385.5, 1275.4, 1173.9, 1649, ...

The PCT values do not follow a normal distribution.


Frequency
0
10
20
30
40
50
60
1000
1200
1400
1600
1800
2000
2200
PCT

The computation of non-parametric tolerance limits

The non-parametric tolerance limits will be based on order statistics and their properties.

Some standard results to be used:

Let X be the random variable having a continuous distribution.

Let $F(x) = P(X \leq x)$ denote the cumulative distribution function (cdf). Then $F(X)$ follows the Uniform(0, 1) distribution.

Let $X_1, X_2, ..., X_n$ be a random sample from a continuous distribution, and let $X_{(1)} < ... < X_{(n)}$ be the order statistics for the sample.

Then $F(X_r)$ follows a Beta$(r; n - r + 1)$ distribution.

Thus the distribution of $F(X_r)$ does not depend on any unknown parameters.

For two order statistics $X_{(r)}$ and $X_{(s)}$, the joint distribution of $F(X_{(r)})$ and $F(X_{(s)})$ does not depend on any unknown parameters.

The following relation exists between the binomial distribution and the beta distribution:

If Y follows the Binomial(n, p) distribution and U follows a Beta($r; n - r + 1$) distribution, then

$$P(Y \geq r) = P(U \leq p),$$

$r = 1, 2, ..., n$.

The above results are used to compute nonparametric tolerance limits that are valid for any continuous distribution.

The tolerance limits are based on order statistics. Which order statistics?

Computation of an upper tolerance limit:

To find r so that $X_{(r)}$ is the upper tolerance limit corresponding to a content p and confidence level $1 - \alpha$.

Let Y follow the Binomial(n, p) distribution.

Let r be the smallest integer for which

$$P(Y \leq r - 1) \geq 1 - \alpha.$$

Then $X_{(r)}$ is the non-parametric upper tolerance limit.

Furthermore, $X_{(n-r+1)}$ is the non-parametric lower tolerance limit.

An example (PCT data):

The data consists of 208 observations. A sample of $n = 50$ observations in the sample:

1615.2, 1799.8, 1742.4, 1624.1, 1682, 1213.5, 1613.1, 1626.7, 1633.4, 1647.8, 1098.8, 1516.4, 1678.4, 1621.6, 1619, 1069, 1508.2, 1050.2, 1738.2, 974.8, 1658.2, 1385.5, 1275.4, 1173.9, 1649, 1649.3, 1597, 1380.7, 1582.9, 1032.9, 1622.7, 1619.1, 1573.5, 1303.4, 1588.7, 1575.8, 1603.8, 1592.4, 1645, 1017.7, 1582.9, 1681.7, 1575.8, 1605.6, 1553.6, 1540.5, 1643.2, 1545.8, 1204.9, 1484.2

Order from smallest to largest:

974.8, 1017.7, 1032.9, 1050.2, 1069, 1098.8, 1173.9, 1204.9, 1213.5, 1275.4, 1303.4, 1380.7, 1385.5, 1484.2, 1508.2, 1516.4, 1540.5, 1545.8, 1553.6, 1573.5, 1575.8, 1575.8, 1582.9, 1582.9, 1588.7, 1592.4, 1597, 1603.8, 1605.6, 1613.1, 1615.2, 1619, 1619.1, 1621.6, 1622.7, 1624.1, 1626.7, 1633.4, 1643.2, 1645, 1647.8, 1649, 1649.3, 1658.2, 1678.4, 1681.7, 1682, 1738.2, 1742.4, 1799.8

We shall take $p = 0.90$ and $1 - \alpha = 0.95$.

Let Y follow the Binomial(n, p) distribution. That is, the Binomial$(50, 0.90)$ distribution.

Let r be the smallest integer for which $P(Y \leq r - 1) \geq 1 - \alpha$. That is $P(Y \leq r - 1) \geq 0.95$.

$P(Y \leq 47) = 0.8882$, $P(Y \leq 48) = 0.9662$

So $r - 1 = 48$, and $r = 49$. The non-parametric upper tolerance limit is $X_{(49)}$, the 49th order statistic, namely 1742.4.

$X_{(n-r+1)} = X_{(2)}$ is the non-parametric lower tolerance limit, which is 1017.7.

Note: The nonparametric tolerance limits are typically conservative

Zimmer, Z., Park, D. and Mathew, T. (2016). Tolerance limits under normal mixtures: Application to the evaluation of nuclear power plant safety and to the assessment of circular error probable. *Computational Statistics & Data Analysis*, 103, 304-315.

Computation of a two-sided tolerance interval:

To find s and t so that $(X_{(s)}, X_{(t)})$ is a two-sided tolerance interval corresponding to a content p and confidence level $1-\alpha$.

Let Y follow the Binomial(n, p) distribution.

Let r be the smallest integer for which

$$P(Y \leq r-1) \geq 1-\alpha.$$

Let s and t be such that $t-s=r$. Then $(X_{(s)}, X_{(t)})$ is a two-sided tolerance interval.

Note: The pair of order statistics providing the two-sided tolerance interval is not unique. It is standard practice to choose $t=n-s+1$ and take $(X_{(s)}, X_{(n-s+1)})$ as the two-sided tolerance interval.

For a given content p and confidence level $1-\alpha$, nonparametric tolerance limits may not exist, unless a minimum sample size requirement is met.

For the upper (or lower) tolerance limit to exist, n must satisfy

$$n \geq \frac{\ln(\alpha)}{\ln(p)}$$

For the two-sided tolerance interval to exist, n must satisfy

$$(n-1)p^n - np^{n-1} + 1 \geq 1-\alpha$$

The following table gives the required minimum sample sizes:

p	Interval Type	$1-\alpha$ 0.90	0.95	0.99
0.90	one-sided	22	29	44
	two-sided	38	46	64
0.95	one-sided	45	59	90
	two-sided	77	93	130
0.99	one-sided	230	299	459
	two-sided	388	473	662

An application:

Liver function data from of Harris and Boyd (1995, Appendix 4.2).

Data on the liver enzyme alanine transaminase (ALT) in u/l taken from 596 healthy male medical students during the years 1987-1991 at the University of Virginia.

Histograms indicate that the data are skewed to the right.

A few observations: 21, 31, 34, 16, 23, 4, 26, 24, 18, 51, 103, 23, ..., 16, 33, 23

Harris, E. K. and Boyd, J. C. (1995). *Statistical Bases of Reference Values in Laboratory Medicine*. Marcel Dekker

To compute a non-parametric reference interval for the ALT using the sample of $n = 596$ observations.

We shall compute a two-sided non-parametric tolerance interval having content and confidence level both equal to 0.95.

Let Y follow the Binomial(n, p) distribution. That is, the Binomial$(596, 0.95)$ distribution.

$P(Y \leq 574) = 0.9461$, $P(Y \leq 575) = 0.9654$.

$r - 1 = 575$. $r = 576$.

Choosing $s = 10$ and $t = 586$ (so that $t - s = r = 576$), the two-sided tolerance interval consists of the 10th and 586th order statistics. The interval is (8, 68).

Tolerance intervals under linear regression models

An example

To compare the breath estimates of blood alcohol concentration (obtained using a breath analyzer) with those determined by a laboratory test.

X: breath estimates obtained using Breathalyzer Model 5000

u: the results of a laboratory test, assumed to be precise (% alcohol concentration in blood)

Data on blood alcohol concentrations (u) and breath estimates (X)

Subject	u	X	Subject	u	X
1	.160	.145	9	.170	.176
2	.170	.156	10	.056	.048
3	.180	.181	11	.111	.092
4	.100	.108	12	.162	.144
5	.170	.180	13	.143	.121
6	.100	.112	14	.079	.065
7	.060	.081	15	.006	.000
8	.100	.104			

Fitted model:

$$X = 0.00135 + 0.958u.$$

The normal probability plot of the residuals is reasonably linear.

Assume normality.

Tolerance interval for the breath estimate distribution: will contain at least a proportion p of the breath estimates with confidence level $1-\alpha$, for a fixed value u of the blood alcohol concentration.

For example, such an interval is required when $u = 0.10$

Consider the simple linear regression model for a response variable X depending on the covariate u:

$$X = \beta_0 + \beta_1 u + e, \quad e \sim N(0, \sigma^2)$$

n independent observations are available:

$$X_i = \beta_0 + \beta_1 u_i + e_i, \quad e_i \sim N(0, \sigma^2), \quad i = 1, ..., n$$

$\hat{\beta}_0, \hat{\beta}_1$: least squares estimates

$\hat{\sigma}^2 = S^2$: Residual mean square

$$\text{For a fixed } u \text{, let } d^2 = \frac{1}{n} + \frac{(u - \overline{u})^2}{\Sigma(u_i - \overline{u})^2}$$

Tolerance factors turn out to be functions of d.

Now consider the response variable X corresponding to the covariate value u.

An upper tolerance limit for X: $(\hat{\beta}_0 + \hat{\beta}_1 u) + kS$

An lower tolerance limit for X: $(\hat{\beta}_0 + \hat{\beta}_1 u) - kS$

$k = k(d)$ is the tolerance factor that satisfies the tolerance interval condition.

$$k(d) = d \times t_{n-1;1-\alpha}(z_p/d),$$

where z_p is the pth percentile of the standard normal distribution.

Example (continued)

X: breath estimates

u: the results of the laboratory test

Suppose we want a lower tolerance limit for X when $u = 0.10$. Choose $p = 0.90$ and $1 - \alpha = 0.95$.

$\hat{\beta}_0 = 0.00135$, $\hat{\beta}_1 = 0.958$, $S = 0.01366$, $d = 0.273643$

$k(d) = 2.1172$

The lower tolerance limit for the breath estimate when $u = 0.10$:
$\hat{\beta}_0 - \hat{\beta}_1 x - k(d)S = 0.068$

We are 95% confident that at least 90% of the breath alcohol estimates exceed 0.068 when the blood alcohol level is 0.10.

To compute a two-sided tolerance interval for X for a fixed u

The interval is taken to be $\hat{\beta}_0 + \hat{\beta}_1 x \pm kS$.

As in the case of the upper and lower tolerance limits, the factor k is a function of d, where

$$d^2 = \frac{1}{n} + \frac{(u - \overline{u})^2}{\Sigma(u_i - \overline{u})^2}$$

An explicit expression is not available for k.

A good approximation is available (next slide)

Example (continued)

To compute a two-sided tolerance interval for the breath alcohol level X when the blood alcohol level is $u = 0.10$. Take $p = 0.90$ and $1 - \alpha = 0.95$.

$n = 15$, $\hat{\beta}_0 = 0.00135$, $\hat{\beta}_1 = 0.958$, $S = 0.01366$, $d^2 = 0.07489$

An approximation is given by

$$k(d) = \frac{(n-2) \times \chi^2_{1;p}(d^2)}{\chi^2_{n-2;\alpha}} = 2.533.$$

The tolerance interval is 0.0971 ± 0.0346

We are 95% confident that when the blood alcohol concentration is 0.10, at least 90% of the breath estimates are in the interval 0.0971 ± 0.0346.

Tolerance intervals can be similarly computed for multiple linear regression models.

Summary of Part 2

Discussed the computation of tolerance limits in a nonparametric setup and under linear regression models.

Illustrated them with real examples.

Part 3: Tolerance regions and reference regions under the multivariate normal distribution, tolerance rectangles, applications in laboratory medicine.

An example:

Assessment of kidney function based on the amounts of urea, uric acid and creatinine in the blood (taken from Albert and Harris (1987).

To construct a trivariate tolerance region for these analytes using data from 284 healthy subjects.

The region can be used for diagnostic purposes. A measurement falling outside the region is indicative of abnormal kidney function.

Taken from the book
Harris, E. K. and Boyd, J. C. (1995). *Statistical Bases of Reference Values in Laboratory Medicine*. Marcel Dekker.

Let $X = (x_1, x_2, x_3)'$ represent the concentrations of urea (mmol/L), uric acid (μ mol/L) and creatinine (μ mol/L) in the blood sample.

Multivariate normality is assumed

Let $\overline{X}$ and S denote the sample mean vector and sample covariance matrix computed based on the data on 284 healthy subjects.

$\overline{X}$ and S, computed from the data, are

$$\overline{X} = \begin{pmatrix} 5.1 \\ 303 \\ 85 \end{pmatrix}, \quad S = \begin{pmatrix} 1.14 & 14.93 & 6.02 \\ 14.93 & 3724.30 & 473.30 \\ 6.02 & 473.30 & 205.50 \end{pmatrix}$$

We shall take $p = 1 - \alpha = 0.95$.

Instead of computing a multivariate tolerance region, why can't we compute separate univariate tolerance intervals for urea, uric acid and creatinine?

Separate univariate tolerance intervals won't take into account the correlations.

The sample correlation matrix is given by

$$\begin{pmatrix} 1.000 & 0.229 & 0.393 \\ 0.229 & 1.000 & 0.541 \\ 0.393 & 0.541 & 1.000 \end{pmatrix}$$

We shall first consider an ellipsoidal tolerance region for $X = (x_1, x_2, x_3)'$. Such a region is given by

$$\{X: (X - \overline{X})' S^{-1} (X - \overline{X}) \leq k\},$$

where k is the tolerance factor to be determined.

The tolerance factor k is computed subject to the condition that the region will include at least 95% of the distribution of the X values, with 95% confidence.

It can be shown that k depends only on n, p and $1 - \alpha$, and does not depend on any unknown parameters.

For various values of n, p and $1 - \alpha$, tables are available for the factor k, for different dimensions.

$$\overline{X} = \begin{pmatrix} 5.1 \\ 303 \\ 85 \end{pmatrix}, \quad S = \begin{pmatrix} 1.14 & 14.93 & 6.02 \\ 14.93 & 3724.30 & 473.30 \\ 6.02 & 473.30 & 205.50 \end{pmatrix}$$

We shall take $p = 1 - \alpha = 0.95$. The tolerance factor $k = 8.68$

The ellipsoidal tolerance region is given by

$$\{X : (X - \overline{X})' S^{-1} (X - \overline{X}) \leq 8.68\}.$$

For $p = 1 - \alpha = 0.95$, the univariate tolerance intervals based on the same data are

Urea: (3.21, 6.99)

Uric Acid: (194.93, 411.07)

Creatinine: (59.62, 110.38)

Recall: $\overline{X} = (5.1, 303, 85)'$.

Consider the two vectors:

$X = (5.1, 500, 85)'$. The uric acid level appears to be abnormal

$X = (8.86, 303, 85)'$. The urea level appears to be abnormal.

The tolerance region is give by

$$\{X: (X - \overline{X})' S^{-1} (X - \overline{X}) \leq 8.68\}.$$

Both of the vectors $X = (5.1, 500, 85)'$ and $X = (8.86, 303, 85)'$ give $(X - \overline{X})' S^{-1} (X - \overline{X}) = 14.74$.

For the first vector, uric acid measurement appears to be abnormal.

For the second vector, the urea measurement appears to be abnormal.

However, the ellipsoidal tolerance region can't distinguish between the two.

In other words, an ellipsoidal tolerance region may not be appropriate to be used as a reference region for diagnostic purposes.

Ellipsoidal tolerance regions can also be constructed for multivariate linear regression models, when multivariate normality holds.

Such regions are relevant in laboratory medicine applications when the reference region depends on covariates such as age, gender, etc.

As an example, such an application is discussed in Mattsson et. al. (2008), where a reference region is constructed for three components of the insulin-like growth factor system in healthy adults.

A. Mattsson, D. Svensson, B. Schuett, K.J. Osterziel, M.B. Ranke (2008). Multidimensional reference regions for IGF-I, IGFBP-2 and IGFBP-3 concentrations in serum of healthy adults, *Growth Hormone and IGF Research*, 18, 506-516.

The three components are:

The serum concentrations (μg/L) of insulin-like growth factor I (S-IGF-I)

Insulin-like growth factor-binding protein 2 (S-IGFBP-2)

Insulin-like growth factor-binding protein 3 (S-IGFBP-3).

Mattsson et. al. (2008) have constructed a reference region using data from 427 healthy individuals.

The three analytes are shown to be associated with elevated levels of growth hormone secretion, and tumor occurrence.

A tolerance region can be used to diagnose an abnormal IGF-IGFBP profile and may be used to assess cancer risk.

The three analytes also depend on the covariates age, gender and BMI.

Based on the data from 427 healthy adults, a regression model was fitted using the covariates: age $-$ 45, (age $-$ 45)^2, gender (male$-$0, female $-$1), and (body mass index(BMI)$-$25).

The need for multivariate reference regions: Another example

Hepatotoxicity: drug-induced injury to the liver; which can eventually cause serious liver damage.

Hepatotoxicity is indicated by elevated levels of the liver enzyme Alanine aminotransferase (ALT) and elevated levels of serum bilirubin.

FDA's *Drug Induced Liver Injury Rank Dataset* consists of 1036 FDA-approved drugs divided into four classes, according to their potential for hepatotoxicity.

From the FDA (2009) guidance document:

"Drug induced liver injury has been the most frequent single cause of safety-related drug marketing withdrawals for the past 50 years ....., continuing to the present ....".

"Hepatotoxicity discovered after approval for marketing also has limited the use of many drugs ....."

U.S. Food and Drug Administration (2009). *Guidance for Industry, Drug-Induced Liver Injury: Premarketing Clinical Evaluation*, U.S. FDA, Rockville, MD.

Hepatotoxicity is assessed using upper reference limits for the ALT and serum bilirubin obtained from a healthy population.

ALT levels beyond thrice the reference limit, and bilirubin levels beyond twice the reference limit indicate hepatotoxicity.

Known as Hy's Law, after the pioneering work of Hy Zimmerman (1999).

While applying Hy's law, the possible correlation between the two variables, ALT and bilirubin levels, is not taken into consideration.

Zimmerman, H. J. (1999). *The Adverse Effects of Drugs and Other Chemicals on the Liver, 2nd Edition*, Lippincott Williams & Wilkins, Philadelphia, PA.

Trost (2006) notes: "... current reference limits are not multivariate, and do not contain any information about the cross-correlations among different tests. In large part, existing reference limits are unscientifically constructed, and cannot be used effectively and consistently in the detection and assessment of abnormality. Furthermore, regulatory agencies rely heavily on these limits for evaluating toxicity."

Trost, D. C. (2006). Multivariate probability-based detection of drug-induced hepatic signals. *Toxicological Reviews*, 25, 37-54.

Trost, D. C. (2015). Hepatotoxicity. In *Statistical Methods for Evaluating Safety in Medical Product Development* (A. L. Gould, Editor), Wiley, pp. 229-270.

Trost, D. C. and Freston, J. W. (2008). Vector analysis to detect hepatotoxicity signals in drug development. *Drug Information Journal*, 42, 27-34.

How to incorporate the correlations while deriving reference regions?

Assuming multivariate normality, can construct ellipsoidal tolerance regions and ellipsoidal central tolerance regions; Dong and Mathew (2015).

Boyd, J. C. (2004). Reference regions for two or more dimensions. *Clinical Chemistry and Laboratory Medicine*, 42, 739-746.

Dong, X. and Mathew, T. (2015). Central tolerance regions and reference regions for multivariate normal populations. *Journal of Multivariate Analysis*, 134, 50-60.

As we have seen, an ellipsoidal reference region is easy to construct under multivariate normality.

However, as we have noticed, an ellipsoidal reference region cannot detect if a particular component of a vector of measurements is extreme.

In a book on laboratory medicine, Strike (1991) notes: "A rather vexing problem is the tendency for multivariate interpretations to misclassify large test profiles that contain just one or two clearly pathological test results. In short, multivariate reference ranges require careful handling" (Section 10.4 in the book).

This calls for the computation of reference regions that are rectangular.

Strike, P. W. (1991). *Statistical Methods in Laboratory Medicine*. Oxford University Press, UK.

An ellipsoidal region is also not appropriate when it is of interest to compute one-sided reference limits, upper or lower.

For the diagnosis of hepatotoxicity, upper reference limits are used. Serum ALT levels beyond thrice the upper reference limit, and bilirubin levels beyond twice the uper reference limit indicate hepatotoxicity.

When reference limits are required for multiple analytes, there are situations where two-sided references limits are required for some of them, but one-sided reference limits for the rest.

For the assessment of kidney function, reference intervals are used for the three analytes: urea, uric acid, and creatinine.

For urea, we only need an upper reference limit, whereas two-sided reference intervals are required for uric acid and creatinine.

Rectangular reference regions under multivariate normality

Let the $q \times 1$ vector $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Write $\mathbf{X} = (X_1, X_2, \cdots, X_q)'$.

Based on a random sample of size n, let $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_q)'$: sample mean vector

$\mathbf{S} = ((S_{ij}))$: sample covariance matrix.

The rectangular region to be constructed is of the form

$$\bar{X}_i \pm k\sqrt{S_{ii}},\quad 1, 2, \cdots, q,$$

where k is the tolerance factor to be determined.

Different types of rectangular tolerance regions:

(i) Simultaneous tolerance intervals: each marginal interval has content p, with a joint confidence level $1-\alpha$.

(ii) Simultaneous central tolerance intervals: each marginal interval is a central tolerance interval having content p, with a joint confidence level $1-\alpha$.

(iii) Tolerance rectangles: the content of the tolerance rectangle is p with confidence level $1-\alpha$.

No recommendation will be made on the criterion to be used in laboratory medicine applications. Needs input from experts in laboratory medicine.

The factor k depends on the unknown population correlations.

A parametric bootstrap approach can be used to estimate the factor k accurately, for each of the three types of rectangular regions.

Not possible to have tables providing the values of k, since it has to be estimated from the data.

However, R codes are available.

Lucagbo, M., Mathew, T. (2023). Rectangular tolerance regions and multivariate normal reference regions in laboratory medicine. *Biometrical Journal, 65:3*.

Example (continued):

X_1 = urea (mmol/L), X_2 = uric acid (μmol/L), and X_3 = creatinine (μmol/L) for the assessment of kidney function.

Data on 284 healthy individuals.

$$\overline{X} = \begin{pmatrix} 5.1 \\ 303 \\ 85 \end{pmatrix}, \quad S = \begin{pmatrix} 1.14 & 14.93 & 6.02 \\ 14.93 & 3724.30 & 473.30 \\ 6.02 & 473.30 & 205.50 \end{pmatrix}$$

Correlation matrix: $\begin{pmatrix} 1.000 & 0.229 & 0.393 \\ 0.229 & 1.000 & 0.541 \\ 0.393 & 0.541 & 1.000 \end{pmatrix}$

Analyte	Tolerance region: Simultaneous	Tolerance region: Simultaneous central	Tolerance region: Rectangular
Urea	(3.17, 7.03)	(3.10, 7.10)	(2.76, 7.44)
Uric Acid	(192.86, 413.14)	(188.92, 417.08)	(169.52, 436.48)
Creatinine	(59.13, 110.87)	(58.20, 111.80)	(53.65, 116.35)

Analyte	Mixed* region
Urea	< 6.68
Uric Acid	(189.99, 416.01)
Creatinine	(58.45, 111.55)

*Upper reference limit for urea, and two-sided reference intervals for uric acid and creatinine.

Multivariate tolerance regions have important applications in laboratory medicine.

The need for such regions are well-documented in the laboratory medicine literature.

Ellipsoidal regions are often computed, even though researchers in laboratory medicine are hesitant to use them.

What is required is a rectangular region that takes into account the correlations.

We have been able to derive such regions by using a parametric bootstrap procedure.

In the literature on laboratory medicine, 95% ellipsoidal prediction regions are often recommended as reference regions.

The same recommendation is also seen in the multivariate case when an ellipsoidal region is desired.

The prediction region criterion is clearly inappropriate, since a prediction region is computed subject to the condition that it includes a future observation with a given confidence level; it is not meant to capture a specified proportion of the population with a high probability.

Other criteria have been proposed and employed: the use of point estimates, confidence intervals on estimated reference limits, and tolerance intervals.

Each involves some trade-off with respect to the specificity and sensitivity of the corresponding diagnostic procedure.

The appropriate criterion for computing reference intervals and regions is still being discussed in the literature.

Liu, W., Bretz, F. and Cortina-Borja, M. (2021). Reference Range: Which Statistical Intervals to Use? *Statistical Methods in Medical Research* 30, 523–534.

Wellek, S. and Jennen-Steinmetz, C. (2022). Reference Ranges: Why Tolerance Intervals Should Not be Used. Comment on Liu, Bretz and Cortina-Borja, Reference Range: Which Statistical Intervals to Use? SMMR, 2021,Vol. 30(2) 523–534. *Statistical Methods in Medical Research* 31, 2255–2256.

Lucagbo, M. D. and Mathew, T. (2023). Rectangular Tolerance Regions and Multivariate Normal Reference Regions in Laboratory Medicine. *Biometrical Journal,* 65:3.

Mathew, T. and Young, D. (2024). A Review of Statistical Reference Regions in Laboratory Medicine: Theory and Computation. Under preparation.

Thank you for attending the short course on statistical tolerance intervals and regions!