Small Sample Asymptotics: Two Applications

Thomas Mathew

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

mathew@umbc.edu

Application 1

Computing **tolerance limits** under the logistic regression model for binary data.

- The problem has been motivated by an application of interest to the U.S. Army.
- Testing of ballistic armor plates for protecting soldiers from projectiles and shrapnel.
- The probability of penetration is a function of covariates such as velocity of the projectile.
- The probability is modeled using the logistic regression model.
- A univariate problem under a discrete model.

Application 2

Testing the **homogeneity of the relative potencies** in a multivariate bioassay problem.

- Data obtained from several independent multivariate bioassays performed at different laboratories or locations.
- The usual slope-ratio or parallel line assay model is assumed.
- A multivariate problem under a continuous model.

Application 1: The ballistic problem

Testing the ballistic resistance of personal body armor.

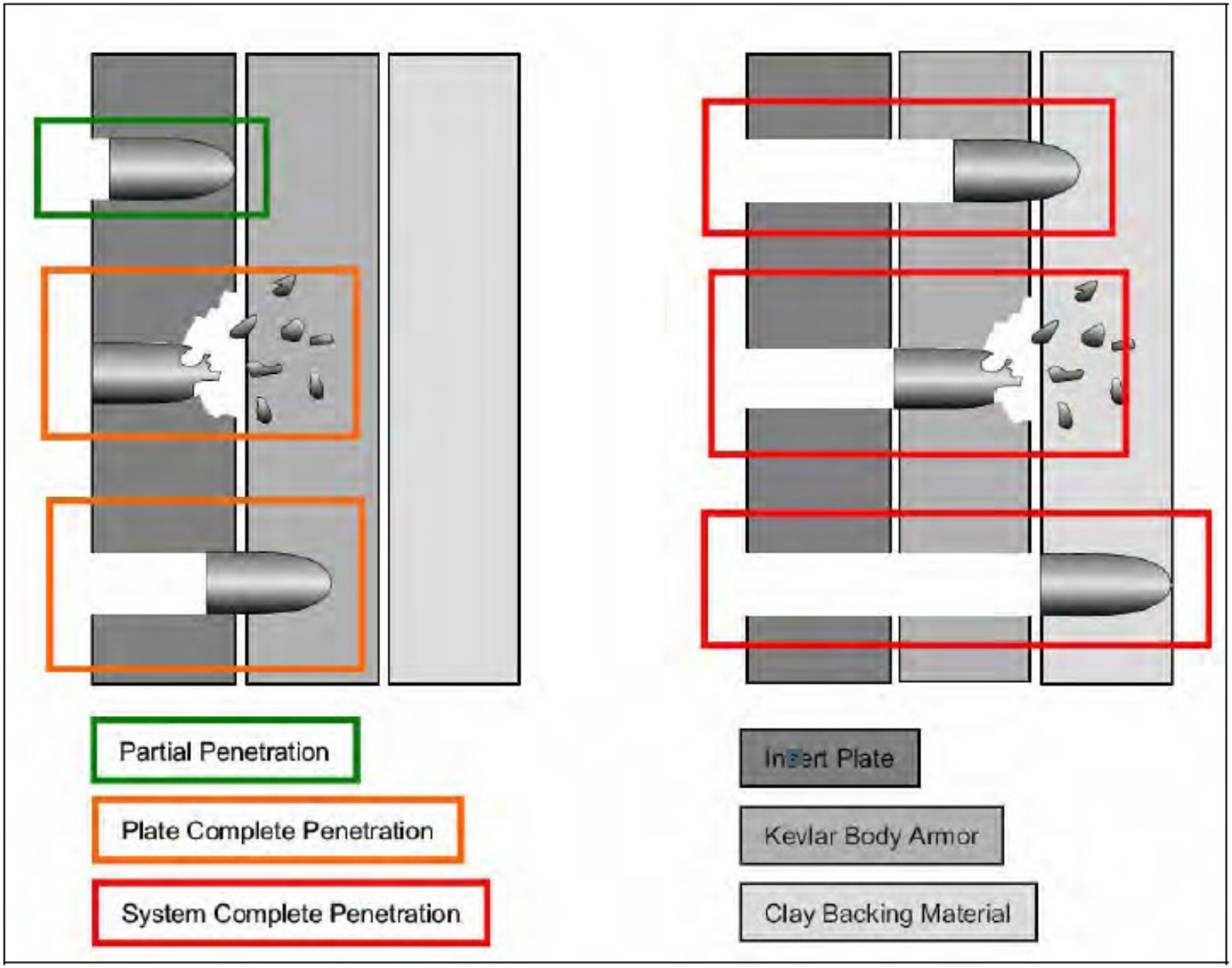
Armor plates are inserted into the soft body armor vests to increase their protective abilities.

The plates come in a variety of sizes, using different materials.

The ballistic performance of these plates are tested against different threats to evaluate their performance.

Covariates such as plate size, material used, and the velocity of the projectile affect their performance.

While testing, a backing material (made of clay) is mounted on the armor plate.


Partial Penetration
Plate Complete Penetration
System Complete Penetration
Insert Plate
Kevlar Body Armor
Clay Backing Material

Two observations are noted after a shot is fired:

(i) if there is penetration, and

(ii) in case of no penetration, the *back face deformation* (BFD), which is the extent of indentation in a backing material.

The BFD is a continuous variable (measured in millimeters).

After performing n tests, the number of penetrations is noted, along with the BFD measurements (in the no penetration cases).

Background information, including the protocols and procedures used in the testing: National Institute of Justice Standards (2008), reports from the National Research Council (2009) and the National Science Academy (2010).

The National Institute of Justice Standards (2008) require that a sample be used to conclude if 80% or more of the BFD values are below 44mm, with a confidence level of 95%.

In the absence of covariates, a $N(\mu, \sigma^2)$ distribution is assumed.

Want to use the sample to conclude if at least 80% of the population distribution is below the threshold value 44mm, with 95% confidence.

What is required is an **upper tolerance limit.**

Tolerance limits and tolerance intervals

The upper tolerance limit is to be computed subject to the condition that at least 80% of the population BFD values are below the limit, with a confidence level of 95%.

Once such an upper tolerance limit is computed, we can verify if it is less than the threshold value 44mm.

A tolerance limit having 80% **content**, and 95% **confidence level.**

A (0.80, 0.95) upper tolerance limit.

Let X_1, X_2, ..., X_n be a random sample from a population.

Write $\boldsymbol{X} = (X_1, X_2, ..., X_n)$.

In order to define a tolerance interval, we need to specify its content, say p, and confidence level, say $1 - \alpha$.

The tolerance interval will be referred to as a p content and $(1 - \alpha)$ confidence, tolerance interval.

A $(p, 1 - \alpha)$ tolerance interval.

The interval will be constructed using the random sample $\boldsymbol{X}$, and is required to contain a proportion p or more of the sampled population, with confidence level $1 - \alpha$.

A $(p, 1-\alpha)$ one-sided tolerance interval of the form $(-\infty, U(\boldsymbol{X})]$ is required to satisfy the condition

$$P_{\boldsymbol{X}}\left\{P_X\left(X \leq U(\boldsymbol{X})\middle|\boldsymbol{X}\right) \geq p\right\} = 1-\alpha$$

The interval $(-\infty, U(\boldsymbol{X})]$ is called a one-sided tolerance interval.

$U(\boldsymbol{X})$ is called a one-sided upper tolerance limit.

$U(\boldsymbol{X})$ is a $100(1-\alpha)\%$ upper confidence limit for the pth percentile of X.

A $(p, 1-\alpha)$ one-sided lower tolerance limit $L(\boldsymbol{X})$ is defined similarly.

$L(\boldsymbol{X})$ is a $100(1-\alpha)\%$ lower confidence limit for the $(1-p)$th percentile of X.

A $(p, 1-\alpha)$ two-sided tolerance interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ satisfies the condition

$$P_{\boldsymbol{X}}\left\{P_X\left(L(\boldsymbol{X}) \leq X \leq U(\boldsymbol{X}) \middle| \boldsymbol{X}\right) \geq p\right\} = 1-\alpha.$$

That is the interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ contains at least a proportion p of the population with confidence $1-\alpha$.

The computation of the two-sided tolerance limit $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ **does not** reduce to the computation of confidence limits for certain percentiles.

Tolerance intervals for a univariate normal population

X_1, X_2, ...,X_n: sample from $N(\mu, \sigma^2)$.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

A $(p, 1-\alpha)$ upper tolerance limit is $\bar{X} + kS$, where

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

The above formula is recommended in the National Institute of Justice Standards (2008), for computing an upper tolerance limit for the BFD distribution.

Upper tolerance limits under logistic regression

To compute an upper tolerance limit on the number of penetrations:

Let $Y = 1$ denote armor plate penetration, $Y = 0$ otherwise.

$\mathbf{x}$: an $s \times 1$ vector of covariates (discrete, as well as continuous)

National Institute of Justice Standards (2008) recommends the logistic regression model for the probability of penetration as a function of the covariates:

$$P(Y = 1) = \pi(\mathbf{x}) = \frac{\exp(\beta_0 + \mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}'\boldsymbol{\beta})}$$

β_0, $\boldsymbol{\beta}$: unknown parameters.

Suppose n independent binary responses $Y_1,\ Y_2,\ ....,\ Y_n$ are available corresponding to the covariate vectors $\mathbf{x}_1,\ \mathbf{x}_2,\ ....,\ \mathbf{x}_n$, respectively.

Consider m future responses, say $Y_{n+1},\ Y_{n+2},\ ....,\ Y_{n+m}$ corresponding to the covariate vector $\mathbf{x}_0$.

Let $W = W(\mathbf{x}_0) = \sum_{i=1}^{m} Y_{n+i}$, so that W is the number of positive responses (i.e., number of penetrations) among the m future responses.

$W \sim$Binomial$(m, \pi(\mathbf{x}_0))$.

An upper tolerance limit for such a W, computed for a fixed $\mathbf{x}_0$, is referred to as a **point-wise upper tolerance limit.**

An upper limit for W that satisfies the tolerance limit condition for all values of the covariate vector (subject to appropriate bounds) is called a **simultaneous upper tolerance limit**.

Write $\boldsymbol{y} = (Y_1, Y_2, ..., Y_n)'$.

An upper tolerance limit for $W(\mathbf{x}_0)$, corresponding to a fixed covariate vector $\mathbf{x}_0$, will be denoted by $U(\mathbf{x}_0, \boldsymbol{y})$, and satisfies:

$$P_{\boldsymbol{y}} \left\{ P_{W(\mathbf{x}_0)} \left[W(\mathbf{x}_0) \leq U(\mathbf{x}_0, \boldsymbol{y}) | \boldsymbol{y} \right] \geq p \right\} \geq 1 - \alpha,$$

where p is the *content* and $1 - \alpha$ is the confidence level.

$U(\mathbf{x}_0, \boldsymbol{y})/m$ is then an upper tolerance limit for the proportion of penetrations among the m future responses.

An upper tolerance limit for the binomial distribution can be obtained from an upper confidence limit for the probability $\pi(\mathbf{x})$.

In the absence of covariates, tolerance limits for the binomial distribution are derived in Hahn and Chandra (1981), Hahn and Meeker (1991, Chapter 6), Zaslavsky (2007), Wang and Tsung (2009) and Krishnamoorthy, Xia, and Xie (2011).

Let $W(\mathbf{x}_0) \sim$Binomial$(m, \pi(\mathbf{x}_0))$.

Let $k_p(\pi(\mathbf{x}_0))$ denote the pth percentile of $W(\mathbf{x}_0)$.

If $\hat{\pi}_U(\mathbf{x}_0)$ is a $100(1-\alpha)\%$ upper confidence limit for $\pi(\mathbf{x}_0)$, then $k_p(\hat{\pi}_U(\mathbf{x}_0))$ is an upper tolerance limit for $W(\mathbf{x}_0)$ with content p and confidence level $1-\alpha$.

The computation of an accurate upper tolerance limit reduces to the computation of an accurate $100(1-\alpha)\%$ upper confidence limit for $\pi(\mathbf{x}_0)$.

The signed log-likelihood ratio test (SLRT) statistic:

Let

$$\psi = \psi(\mathbf{x}_0) = \text{logit}(\pi(\mathbf{x}_0)) = \beta_0 + \mathbf{x}_0'\boldsymbol{\beta},$$

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}'\boldsymbol{\beta})} = \frac{\exp(\psi + (\mathbf{x} - \mathbf{x}_0)'\boldsymbol{\beta})}{1 + \exp(\psi + (\mathbf{x} - \mathbf{x}_0)'\boldsymbol{\beta})}.$$

Let $\boldsymbol{\theta}' = (\psi, \boldsymbol{\beta}')$

$\ell(\boldsymbol{\theta})$: the log-likelihood function

$\hat{\boldsymbol{\theta}}$: the MLE of $\boldsymbol{\theta}$

$\hat{\boldsymbol{\theta}}_\psi$: the constrained MLE of $\boldsymbol{\theta}$, keeping ψ fixed.

The signed log-likelihood ratio test statistic is given by

$$r(\psi) = \text{sign}(\hat{\psi} - \psi) \left[2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_\psi)\} \right]^{1/2},$$

where sign(x) is $+1$ or -1, depending on whether x is positive or negative, respectively.

$r(\psi)$ has an asymptotic standard normal distribution (($O(n^{-1/2})$).

Modified Signed Log-Likelihood Ratio Test (MSLRT) Statistic:

Suggested by Davison, Fraser and Reid (2006) in order to achieve improved small sample performance.

Applicable to general models for discrete data.

Improved version:

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \ln\left(\frac{q(\psi)}{r(\psi)}\right).$$

Involves a factor $q(\psi)$

$r(\psi)$ has an asymptotic standard normal distribution (($O(n^{-1})$).

Derivation of $q(\psi)$:

$$\begin{aligned}
\boldsymbol{v}_i &= \left.\frac{d\pi(\mathbf{x}_i)}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\exp(\hat{\psi} + (\mathbf{x}_i - \mathbf{x}_0)'\hat{\boldsymbol{\beta}})}{\left[1 + \exp(\hat{\psi} + (\mathbf{x}_i - \mathbf{x}_0)'\hat{\boldsymbol{\beta}})\right]^2}(1,\ (\mathbf{x}_i - \mathbf{x}_0)')' \\
\phi(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left[\psi + (\mathbf{x}_i - \mathbf{x}_0)'\boldsymbol{\beta})\right] \boldsymbol{v}_i.
\end{aligned}$$

$$\begin{aligned}
\frac{d\psi}{d\boldsymbol{\phi}} &= \left[\frac{d\boldsymbol{\phi}}{d\boldsymbol{\theta}}\right]^{-1} \frac{d\psi}{d\boldsymbol{\theta}} = \left[\sum_{i=1}^{n} \boldsymbol{v}_i \left\{1, (\mathbf{x}_i - \mathbf{x}_0)'\right\}\right]^{-1} (1, 0, 0, ..., 0)' \\
\chi(\boldsymbol{\theta}) &= \left.\frac{[d\psi/d\boldsymbol{\phi}]'}{||d\psi/d\boldsymbol{\phi}||}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_\psi} \times \boldsymbol{\phi}(\boldsymbol{\theta}).
\end{aligned}$$

$J_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta})$: observed information matrix

$\phi_{\boldsymbol{\theta}}(\boldsymbol{\theta})$: derivative of $\phi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$

$J_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta})$: the $(\boldsymbol{\beta}\boldsymbol{\beta})-$block of $J_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta})$

$\phi_{\boldsymbol{\beta}}(\boldsymbol{\theta})$: the derivative of $\phi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$.

$$q(\psi) = |\chi(\hat{\boldsymbol{\theta}}) - \chi(\hat{\boldsymbol{\theta}}_\psi)| \times \left\{ \frac{|J_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})| \times |\boldsymbol{\phi}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})|^{-2}}{|J_{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_\psi)| \times |\boldsymbol{\phi}'_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_\psi)\boldsymbol{\phi}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_\psi)|^{-1}} \right\}^{1/2}.$$

The MSLRT statistic is

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \ln\left(\frac{q(\psi)}{r(\psi)}\right).$$

$r(\psi)$ has an asymptotic standard normal distribution ($(O(n^{-1}))$)

Explicit expressions are available for all the quantities, except the MLE $\hat{\boldsymbol{\theta}}$ and the constrained MLE $\hat{\boldsymbol{\theta}}_\psi$.

Effort required to compute $r^*(\psi)$ is the same as that required to compute $r(\psi)$.

Performance of $r(\psi)$ and $r^*(\psi)$:

Numerical results on the coverage probabilities of upper confidence limits for ψ.

Consider

$$\text{logit}(\pi(x)) = 6.907 - 14.5x.$$

Nominal level $= 95\%$.

	$x = 0.3$		$x = 0.5$		$x = 0.7$		$x = 0.9$	
n	$r(\psi)$	$r^*(\psi)$	$r(\psi)$	$r^*(\psi)$	$r(\psi)$	$r^*(\psi)$	$r(\psi)$	$r^*(\psi)$
30	0.965	0.820	0.896	0.938	0.840	1.000	0.844	1.000
40	0.968	0.927	0.910	0.960	0.906	0.983	0.898	0.991
50	0.966	0.951	0.932	0.953	0.910	0.964	0.907	0.963
60	0.963	0.950	0.935	0.954	0.909	0.957	0.900	0.951
70	0.963	0.950	0.937	0.951	0.909	0.949	0.907	0.947
80	0.966	0.951	0.934	0.948	0.914	0.953	0.913	0.950
90	0.963	0.951	0.940	0.952	0.921	0.952	0.915	0.948
100	0.964	0.950	0.941	0.951	0.922	0.952	0.921	0.951

If the model involves several covariates and/or interactions, larger sample sizes are required to guarantee satisfactory performance of $r^*(\psi)$.

Example:

Armor plate penetration data obtained from $n = 447$ shots taken at various velocities using a total of 149 plates, with three shots per plate.

Among the 149 plates, there were 52 extra-large (XL), 46 extra-small (XS), and 41 medium (MD) plates.

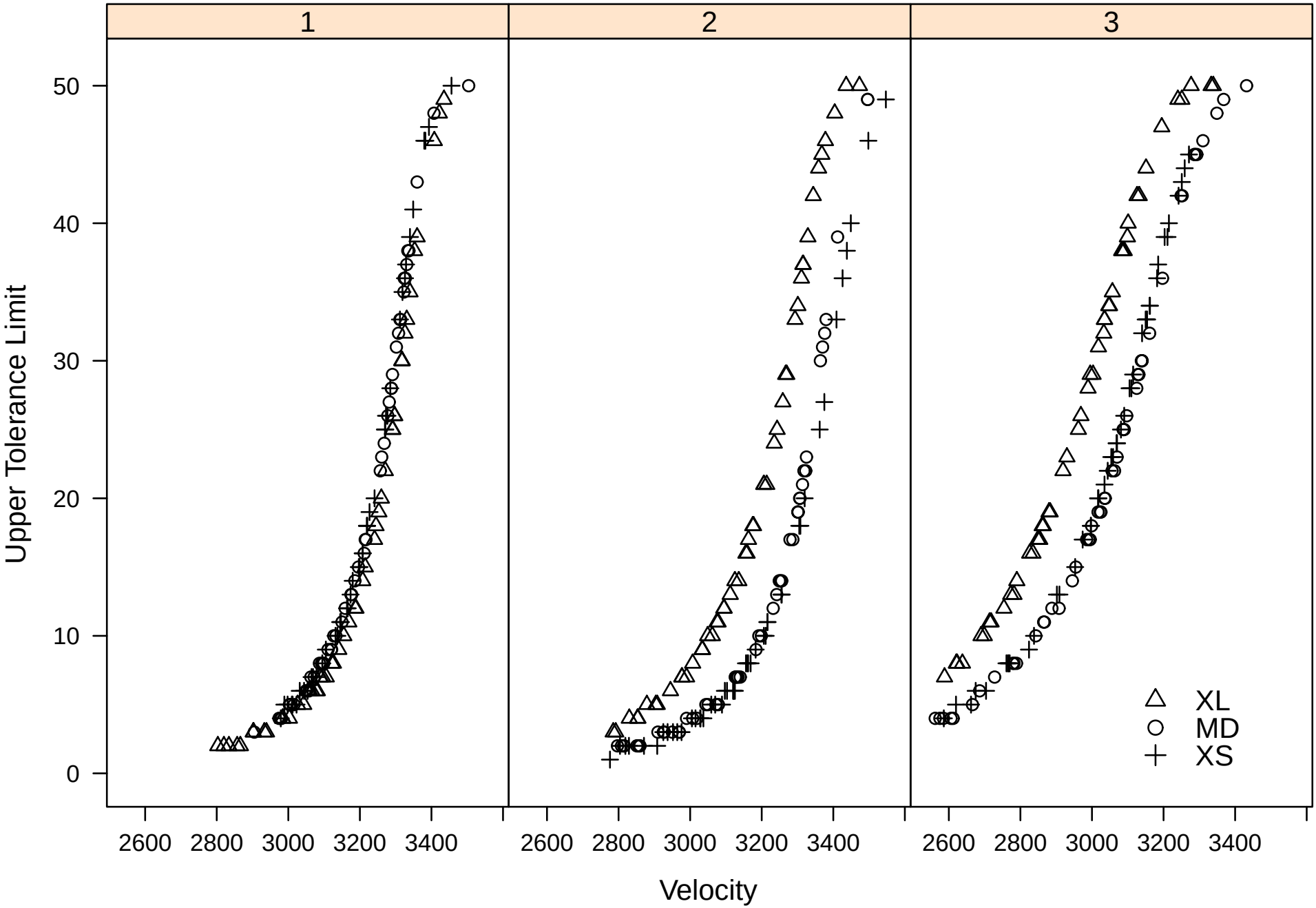
The plate size is considered a factor due to possible lot to lot differences as well as due to the overall surface area differences.

The shot number was included as a discrete covariate, with three categories (since there were three shots per plate).

The logistic regression model also included an interaction between shot number and velocity, and the interaction between plate size and shot number.

To compute upper tolerance limits for the number of penetrations among $m = 50$ plates (representing a future lot size), at various velocity values for each plate size and shot number.

Used 95% content and 95% confidence level.


1
2
3
Upper Tolerance Limit
0
10
20
30
40
50
2600
2800
3000
3200
3400
Velocity
XL
MD
XS

At low velocity values, the number of penetrations is expected to be low, which is to be anticipated.

The plot can be used to draw conclusions regarding the velocity values at which the number of penetrations remain low.

For the second and third shots, the extra-large plate size results in a larger upper tolerance limit compared to the other two plate sizes, at every fixed velocity value.

Application 2. Testing homogeneity in multivariate bioassays

Pain relief data from a dental study reported in Laska et. al. (1983).

Standard treatment (S): 500mg of acetaminophen given as 1, 2, or 3 tablets (dosage levels of the standard treatment are 500mg, 1000 mg and 1500 mg).

Test treatment (T): same three dosage levels along with 65 mg of caffeine.

Two pain intensity scores are recorded on each patient, so that we have bivariate data.

$(y_{S1}, y_{S2})'$: Data on the standard treatment

$(y_{T1}, y_{T2})'$: Data on the test treatment

Data obtained from three locations.

Location 1			Location 2			Location 3		
dose	y_{S1}	y_{S2}	dose	y_{S1}	y_{S2}	dose	y_{S1}	y_{S2}
500	6.5	0.5	500	0.0	0.0	500	7.5	11.5
500	0.0	0.0	500	6.0	9.5	500	8.0	12.0
500	7.0	10.0	500	8.0	11.5	500	4.0	8.0
500	1.0	3.5	1000	3.0	5.0	500	4.0	7.0
1000	8.5	12.0	1000	5.0	8.5	1000	2.0	3.5
1000	10.0	14.0	1500	4.0	7.5	1000	8.0	12.0
1000	6.5	7.5	1500	4.5	7.5	1000	1.0	1.0
1000	1.0	4.0	1500	6.0	9.5	1000	8.0	12.0
1000	3.5	4.5				1500	3.5	7.0
1500	7.0	10.5				1500	1.0	1.0
1500	3.0	6.0				1500	8.0	12.0
1500	6.0	8.0				1500	4.0	8.0
1500	3.0	5.0						
1500	6.5	10.0						
1500	3.5	6.5						

Location 1			Location 2			Location 3		
dose	y_{T1}	y_{T2}	dose	y_{T1}	y_{T2}	dose	y_{T1}	y_{T2}
500	3.5	6.0	500	5.0	8.5	500	6.0	9.0
500	5.0	8.0	500	6.0	9.5	500	7.5	11.5
500	1.0	4.0	1000	5.0	8.0	500	8.0	12.0
1000	10.5	14.5	1000	6.0	9.5	500	6.5	10.0
1000	6.0	7.5	1000	4.5	6.5	1000	8.0	12.0
1500	3.5	8.0	1500	7.0	10.5	1000	7.5	11.5
1500	10.0	14.0	1500	7.0	11.0	1000	0.0	0.0
			1500	5.0	8.5	1500	6.0	10.0
			1500	4.0	4.0	1500	7.5	11.5
						1500	8.0	12.0
						1500	7.5	11.5

Data has been analyzed using a parallel-line assay model.

First consider a single bioassay.

Standard treatment S is assayed at m_S doses $x_{S1}, \ldots, x_{Sm_S}$, with n_{Sj} replicates for dose x_{Sj}.

Test treatment T is assayed at m_T doses $x_{T1}, \ldots, x_{Tm_T}$, with $n_{Tj'}$ replicates for dose $x_{Tj'}$.

$p \times 1$ vector of responses:

$\boldsymbol{y}_{Sjk}$ $(k = 1, 2, \ldots., n_{Sj})$ at dose level x_{Sj}

$\boldsymbol{y}_{Tj'k'}$ $(k' = 1, 2, \ldots., n_{Tj'})$ at dose level $x_{Tj'}$.

Parallel-line assay model:

$$\begin{aligned} \boldsymbol{y}_{Sjk} &= \boldsymbol{\beta}_0 + \boldsymbol{\beta} x_{Sj} + \boldsymbol{\epsilon}_{Sjk} \\ \boldsymbol{y}_{Tj'k'} &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}(\mu + x_{Tj'}) + \boldsymbol{\epsilon}_{Tj'k'}. \end{aligned}$$

$\boldsymbol{\beta}_0$, $\boldsymbol{\beta}$: unknown $p \times 1$ vectors

μ: unknown log-relative potency

$\boldsymbol{\epsilon}_{Sjk}$, $\boldsymbol{\epsilon}_{Tj'k'}$: independent $\sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Typically, x_{Sj} and $x_{Tj'}$ are log doses.

A canonical form

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \boldsymbol{\beta} \\ \mu\boldsymbol{\beta} \end{pmatrix}, D \otimes \boldsymbol{\Sigma}\right],$$

$$\boldsymbol{M} \sim W_p\left(\frac{\boldsymbol{\Sigma}}{m}, m\right), \text{ where } D \text{ is a known matrix.}$$

When there are k independent multivariate bioassays,

$$\begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \boldsymbol{\beta}_i \\ \mu_i\boldsymbol{\beta}_i \end{pmatrix}, D_i \otimes \boldsymbol{\Sigma}_i\right],$$

$$\boldsymbol{M}_i \sim W_p\left(\frac{\boldsymbol{\Sigma}_i}{m_i}, m_i\right).$$

D_i's: known matrices,

$\boldsymbol{\beta}_i$'s: unknown $p \times 1$ vectors,

μ_i's: unknown log-relative potencies,

$\boldsymbol{\Sigma}_i$'s: unknown $p \times p$ positive definite matrices, $i = 1, 2, \ldots, k$.

Two problems of interest:

(i) Develop a test for the null hypothesis $\mu_1 = \mu_2 = ... = \mu_k$ (the relative potencies are homogeneous across the different studies).

(ii) Assuming that $\mu_1 = \mu_2 = ... = \mu_k = \mu$, derive a confidence interval for μ.

Here we address problem (i).

Likelihood based methods assuming that the $\boldsymbol{\Sigma}_i$'s are equal:

Volund (1982)

Meisner, Kushner and Laska (1986)

Srivastava (1986)

Hanusz (1995)

Chen, Carter, Hubert and Kim (1999).

Higher order inference

$\boldsymbol{y}$: $n \times 1$ vector of observations whose distribution depends on a $d \times 1$ parameter vector $\boldsymbol{\theta}$.

$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \boldsymbol{y})$: the log-likelihood function.

$\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda}')'$, where ψ is an $b \times 1$ parameter of interest and $\boldsymbol{\lambda}$ is a nuisance parameter.

$\hat{\boldsymbol{\theta}} = (\hat{\psi}, \hat{\boldsymbol{\lambda}}')'$: MLE

$\hat{\boldsymbol{\lambda}}_{\psi}$: constrained MLE of $\boldsymbol{\lambda}$ for a fixed ψ,

$\hat{\boldsymbol{\theta}}_{\psi} = (\psi, \hat{\boldsymbol{\lambda}}_{\psi}')'$.

Let

$$w(\psi) = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})\},$$

so that $w(\boldsymbol{\psi})$ has an asymptotic chisquare distribution with df $= b$ (LRT).

Tail area approximation has first order accuracy.

Modifications due to Skovgaard (2001) to guarantee second order accuracy.

$\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta})$: score function

$\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1)$ and $\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}_2)$: the score function $\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, evaluated at two parameter points $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

Define the $d \times d$ matrix $\mathbf{S}$ and the $d \times 1$ vector $\mathbf{q}$ as

$$\begin{aligned} \mathbf{S} &= \text{Cov}_{\boldsymbol{\theta}_1}\left\{\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1), \ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}_2)\right\} | \boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_\psi \\ \mathbf{q} &= \text{Cov}_{\boldsymbol{\theta}_1}\left\{\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1), \ell(\boldsymbol{\theta}_1) - \ell(\boldsymbol{\theta}_2)\right\} | \boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_\psi. \end{aligned}$$

The covariances given above are first obtained at $\boldsymbol{\theta} = \boldsymbol{\theta}_1$, and then evaluated at $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_\psi$.

$J(\boldsymbol{\theta})$ and $I(\boldsymbol{\theta})$: the observed and expected information matrices,

$J_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\boldsymbol{\theta})$: the $\boldsymbol{\lambda}\boldsymbol{\lambda}-$block of $J(\boldsymbol{\theta})$.

Let

$$\begin{aligned}\gamma(\boldsymbol{\psi}) &= \frac{|I(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})|^{1/2}|I(\hat{\boldsymbol{\theta}})|^{1/2}}{|\boldsymbol{S}||J_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})^{-1/2}|}\left|[I(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})\boldsymbol{S}^{-1}J(\hat{\boldsymbol{\theta}})I(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{S}]_{\boldsymbol{\lambda}\boldsymbol{\lambda}}\right|^{-1/2} \\ &\quad \times \frac{\left\{\ell'_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})\boldsymbol{S}^{-1}I(\hat{\boldsymbol{\theta}})J(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{S}I(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})^{-1}\ell_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})\right\}^{s/2}}{w(\boldsymbol{\psi})^{b/2-1}\ \ell'_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\psi}})\boldsymbol{S}^{-1}\boldsymbol{q}}\end{aligned}$$

Skovgaard (2001) has proposed two modified statistics:

$$\begin{aligned}w^*(\boldsymbol{\psi}) &= w\left(1-\frac{1}{w}\ln\gamma\right)^2, \\ w^{**}(\boldsymbol{\psi}) &= w-2\ln\gamma\end{aligned}$$

The statistics w^* and w^{**} have asymptotic chisquare distributions with df $= b$, where b is the dimension of $\boldsymbol{\psi}$. The relative error is $O(n^{-1})$.

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics* **28**, 3–32.

Testing the Homogeneity of the Relative Potencies

Canonical form for k independent multivariate bioassays:

$$\begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \boldsymbol{\beta}_i \\ \mu_i \boldsymbol{\beta}_i \end{pmatrix}, D_i \otimes \boldsymbol{\Sigma}_i\right],$$

$$\boldsymbol{M}_i \sim W_p\left(\frac{\boldsymbol{\Sigma}_i}{m_i}, m_i\right).$$

To test $H_0 : \mu_1 = \mu_2 = ... = \mu_k$.

Take $\boldsymbol{\psi}$ to be $k-1$ orthogonal contrasts among the μ_i's, so that we have H_0: $\boldsymbol{\psi} = 0$.

To compare the likelihood ratio test based the chisquare distribution of $w(\psi)$, and the higher order modifications $w^*(\psi)$ and $w^{**}(\psi)$.

The higher order modifications have been implemented after obtaining the score vector, the observed and expected information matrices, the matrix $\boldsymbol{S}$ and the vector $\boldsymbol{q}$.

These quantities can be obtained explicitly.

Effort required to compute $w^*(\psi)$ and $w^{**}(\psi)$ is the same as that required to compute $w(\psi)$.

Type I error probabilities of the tests based on $w(\boldsymbol{\psi})$ (the LRT), $w^*(\boldsymbol{\psi})$ and $w^{**}(\boldsymbol{\psi})$, for a 5% significance level when (m_1, m_2, m_3)=(5, 5, 5); bivariate case

δ_1	δ_2	δ_3	w	w^*	w^{**}
–0.9	–0.9	–0.9	0.2134	0.0657	0.0471
		0.9	0.2130	0.0630	0.0443
	0.1	–0.9	0.2148	0.0671	0.0481
		0.9	0.2153	0.0619	0.0431
	0.9	–0.9	0.2102	0.0630	0.0448
		0.9	0.2128	0.0628	0.0442
0.9	–0.9	–0.9	0.2212	0.0690	0.0504
		0.9	0.2215	0.0664	0.0488
	0.1	–0.9	0.2244	0.0655	0.0460
		0.9	0.2175	0.0651	0.0476
	0.9	–0.9	0.2263	0.0706	0.0494
		0.9	0.2222	0.0677	0.0484

Type I error probabilities of the tests based on $w(\boldsymbol{\psi})$ (the LRT), $w^*(\boldsymbol{\psi})$ and $w^{**}(\boldsymbol{\psi})$, for a 5% significance level when (m_1, m_2, m_3)=(20, 20, 20); bivariate case.

δ_1	δ_2	δ_3	w	w^*	w^{**}
–0.9	–0.9	–0.9	0.0751	0.0460	0.0452
		0.9	0.0783	0.0486	0.0474
	0.1	–0.9	0.0775	0.0517	0.0509
		0.9	0.0796	0.0511	0.0504
	0.9	–0.9	0.0815	0.0516	0.0505
		0.9	0.0783	0.0475	0.0465
0.9	–0.9	–0.9	0.0871	0.0539	0.0527
		0.9	0.0813	0.0501	0.0484
	0.1	–0.9	0.0833	0.0527	0.0516
		0.9	0.0793	0.0520	0.0510
	0.9	–0.9	0.0777	0.0484	0.0477
		0.9	0.0819	0.0519	0.0505

Example (continued)

Pain relief data from a dental study reported in Laska et. al. (1983).

Standard treatment: 500mg of acetaminophen given as 1, 2, or 3 tablets (dosage levels of the standard treatment are 500mg, 1000 mg and 1500 mg).

Test treatment: same three dosage levels along with 65 mg of caffeine.

Two pain intensity scores are recorded on each patient, so that we have bivariate data.

$(y_{S1}, y_{S2})'$: Data on the standard treatment

$(y_{T1}, y_{T2})'$: Data on the test treatment

Data obtained from three locations.

Location 1			Location 2			Location 3		
dose	y_{S1}	y_{S2}	dose	y_{S1}	y_{S2}	dose	y_{S1}	y_{S2}
500	6.5	0.5	500	0.0	0.0	500	7.5	11.5
500	0.0	0.0	500	6.0	9.5	500	8.0	12.0
500	7.0	10.0	500	8.0	11.5	500	4.0	8.0
500	1.0	3.5	1000	3.0	5.0	500	4.0	7.0
1000	8.5	12.0	1000	5.0	8.5	1000	2.0	3.5
1000	10.0	14.0	1500	4.0	7.5	1000	8.0	12.0
1000	6.5	7.5	1500	4.5	7.5	1000	1.0	1.0
1000	1.0	4.0	1500	6.0	9.5	1000	8.0	12.0
1000	3.5	4.5				1500	3.5	7.0
1500	7.0	10.5				1500	1.0	1.0
1500	3.0	6.0				1500	8.0	12.0
1500	6.0	8.0				1500	4.0	8.0
1500	3.0	5.0						
1500	6.5	10.0						
1500	3.5	6.5						

Location 1			Location 2			Location 3		
dose	y_{T1}	y_{T2}	dose	y_{T1}	y_{T2}	dose	y_{T1}	y_{T2}
500	3.5	6.0	500	5.0	8.5	500	6.0	9.0
500	5.0	8.0	500	6.0	9.5	500	7.5	11.5
500	1.0	4.0	1000	5.0	8.0	500	8.0	12.0
1000	10.5	14.5	1000	6.0	9.5	500	6.5	10.0
1000	6.0	7.5	1000	4.5	6.5	1000	8.0	12.0
1500	3.5	8.0	1500	7.0	10.5	1000	7.5	11.5
1500	10.0	14.0	1500	7.0	11.0	1000	0.0	0.0
			1500	5.0	8.5	1500	6.0	10.0
			1500	4.0	4.0	1500	7.5	11.5
						1500	8.0	12.0
						1500	7.5	11.5

Data has been analyzed using a parallel-line assay model.

$m_1 = 22, \;\; m_2 = 13, \;\; m_3 = 19$

The hypothesis of equality of the covariance matrices at the three locations is rejected (p-value less than 0.0001).

For testing the homogeneity of the relative potencies, the different test statistics have values $w(\boldsymbol{\psi}) = 4.164$, $w^*(\boldsymbol{\psi}) = 1.963$ and $w^{**}(\boldsymbol{\psi}) = 1.554$.

The corresponding p-values are 0.1246, 0.3747 and 0.4597.

Thus we accept the null hypothesis of a common relative potency.

Talk based on two articles:

Zimmer, Z., Park, D. and Mathew, T. (2014). Point-wise and simultaneous tolerance limits under logistic regression. *Technometrics*, **56**, 282-290.

Sharma, G., Mathew, T. and Bebu, I. (2014). Combining multivariate bioassays: Accurate inference using small sample asymptotics. *Scandinavian Journal of Statistics*, **41**, 152-166.

R-codes are available

REFERENCES

Barndorff-Nielsen, O. E. (1991). Modified signed log-likelihood ratio. *Biometrika*, 78, 557 – 563.

Brazzale, A. R., Davison, A. C. and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*, Cambridge: Cambridge University Press.

Chen, D. G., Carter, E. M., Hubert, J. J. and Kim, P. T. (1999). Empirical Bayes estimation for combinations of multivariate bioassays. *Biometrics*, 55, 1038–1043.

Davison, A.C., Fraser, D. A. S. and Reid, N. (2006). Improved likelihood inference for discrete data. *Journal of the Royal Statistical Society, Series B*, 68, 495-508.

Davison, A.C., Fraser, D. A. S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of the American Statistical Association*, 109, 302-314.

DiCiccio, T. J., Martin, M. A. and Stern, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *The Canadian Journal of Statistics*, 67, 67-76.

Hahn, G. J. and Chandra, R. (1981). Tolerance intervals for Poisson and binomial variables., *Journal of Quality Technology*, 13, 100-110.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide to Practitioners*, New York: John Wiley.

Hanusz, Z. (1995). Relative potency of two preparations in two-way elimination of heterogeneity designs with multivariate responses. *Biometrics*, 51, 1133–1139.

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

Krishnamoorthy, K., Xia, Y. and Xie, F. (2011). A simple approximate procedure for constructing tolerance intervals for binomial and Poisson distributions. *Communications in Statistics -Theory and Methods*, 40, 2443-2458.

Laska, E. M., Sunshine, A., Zighelbiom, I., Roure, C., Marrero, I., Wanderling, J. and Olson, N. (1983). Effect of caffeine on acetaminophen analgesia. *Clinical and Pharmocological Therapeutics*, 11, 498–427.

Meisner, M., Kushner, H. B. and Laska, E. M. (1986). Combining multivariate bioassays. *Biometrics* 42, 421–427.

National Institute of Justice (2008). Ballistic Resistance of Body Armor. NIJ Standard 0101.06. Washington, D.C.: National Institute of Justice.

National Research Council (2009). Phase I Report on Review of the Testing of Body Armor Materials for Use by the U.S. Army. Washington, D.C.: The National Academies Press.

National Science Academy (2010). Testing of Body Armor Materials for Use by the U.S. Army–Phase II: Letter Report. Washington, D.C.: The National Academies Press.

Sharma, G. and Mathew, T. (2012). One-sided and two-sided tolerance intervals in general mixed and random effects models using small sample asymptotics. *Journal of the American Statistical Association*, 107, 258-267.

Skovgaard, I. M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, 2, 145–165.

Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28, 3–32.

Srivastava, M. S. (1986). Multivariate bioassay, combination of bioassays and Fiellers theorem. *Biometrics*, 42, 131– 141.

Volund, A. (1982). Combination of multivariate bioassay. *Biometrics*, 38, 181–190.

Wang, H. and Tsung, F. (2009). Tolerance intervals with improved coverage probabilities for binomial and Poisson variables. *Technometrics*, 51, 25-33.

Zaslavsky, B. G. (2007). Calculation of tolerance limits and sample size determination for clinical trials with dichotomous outcomes. *Journal of Biopharmaceutical Statistics*, 17, 481-491.