

**Statistical Methods for Interlaboratory Studies:
A Historical Perspective**

Mark G. Vangel
Statistical Engineering Division
National Institute of
Standards and Technology
100 Bureau Drive, Stop 8980
Gaithersburg, MD 20899-8980

Los Alamos National Laboratory
October 24, 20000

Outline

- Introduce interlaboratory studies, with motivation and examples.
- Review influential work of W.J. Youden and John Mandel
 - Youden plots, Ranking
 - Row-Linear Models, Mandel-Paule Algorithm
- Present some recent work
 - Generalize Youden plot
 - Mandel-Paule procedure as approximate REML
 - ML and profile likelihood analysis.
 - Bayesian inference, including ranking

Interlaboratory Studies: The Scenario

- Each of p laboratories makes repeated measurements of m quantities (perhaps corresponding to different concentrations of a chemical analyte).
- The number of measurements made can differ among the laboratories.
- The measurement variability may depend on the material being measured (perhaps as an increasing function of concentration or level).
- The within-laboratory variabilities may differ (often, though, they are assumed to be equal).

**Interlaboratory Studies:
Some questions**

- How should one estimate ‘consensus’ values of the quantities measured?
- What is the between-laboratory variability (*reproducibility*)?
- What is the within-laboratory variability (*repeatability*)? How do they compare?
- How should we look for outliers?
- Which labs perform adequately; which have problems?

Why Interlaboratory Studies?

- Interlaboratory studies are primarily performed for one of two reasons:
 1. Validating a measurement method or standard material
 2. Assessing the proficiency of measurement laboratories.

Examples of Both Types of Studies

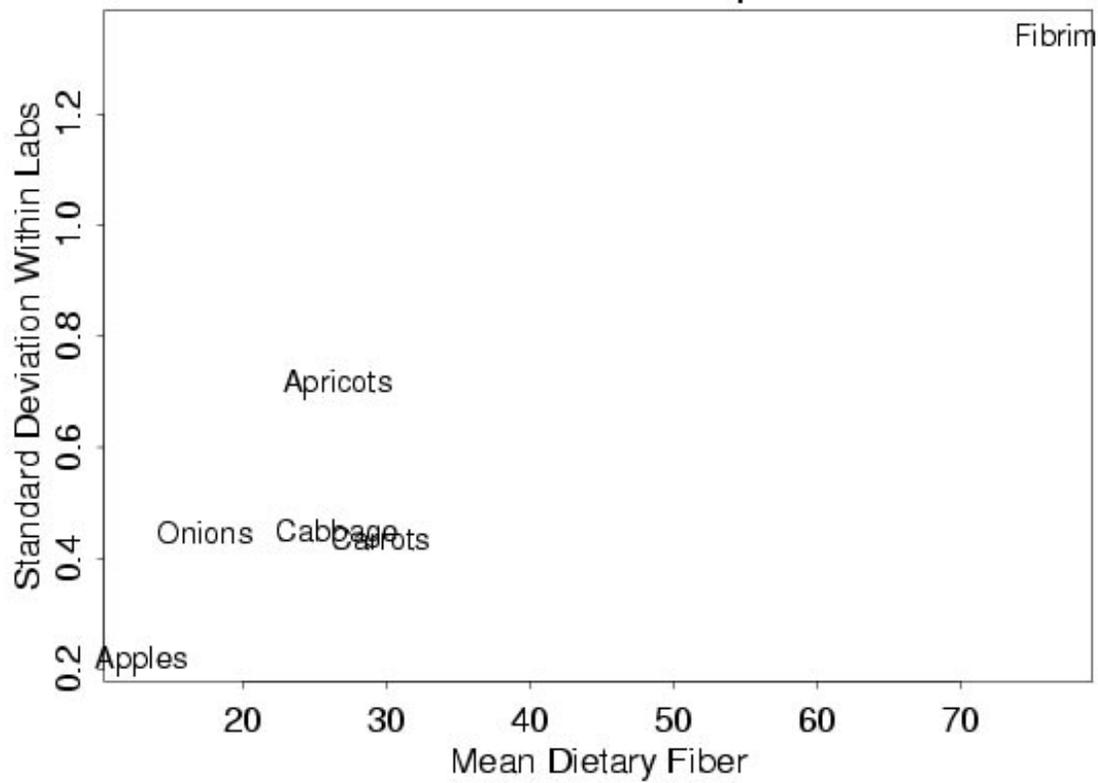
- An enzymatic-gravimetric method is developed for measuring the dietary fiber in foods. Standardized samples of foods are prepared, and distributed to various testing laboratories, who measure the concentrations using the proposed method.
- The National Research Council of Canada and NOAA together conduct interlaboratory comparisons to evaluate the proficiency of test laboratories at determining concentrations of trace elements in marine biological tissues. Homogeneous materials (e.g. oyster tissue, marine sediments) are distributed among various laboratories, who return data on several trace elements (e.g., arsenic).

Evaluating an Analytical Method for Dietary Fiber
Li and Cardozo (1994)
J. Of AOAC Int., 77, p. 689

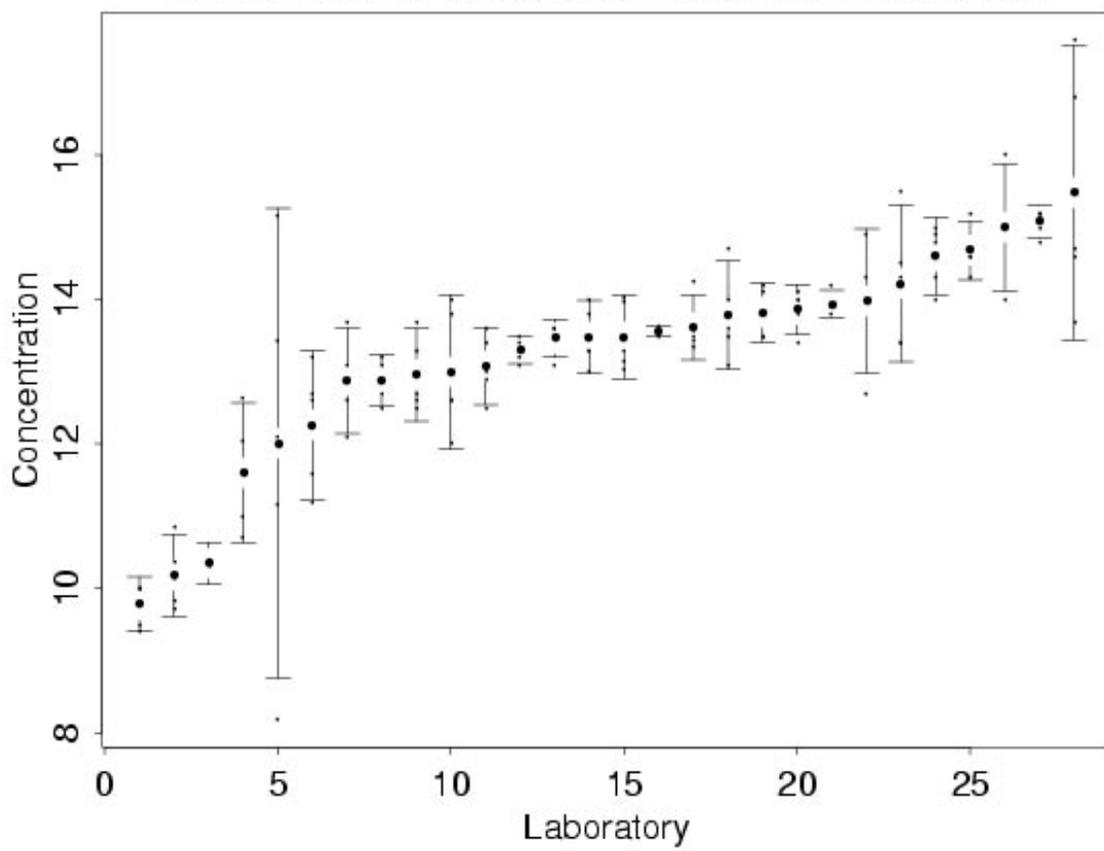
Nine labs each measures fiber in six foods, in *blind duplicates*.

Sample	Laboratory			
	1	2	...	9
Apples	12.44	12.87	...	12.08
	12.48	13.20	...	12.38
Apricots	25.05	27.16	...	25.31
	25.58	26.29	...	25.43
⋮	⋮	⋮	...	⋮
FIBRIM	74.07	76.55	...	73.96
	75.01	78.36	...	74.24

Dietary Fiber Data: Within-Lab. Stand. Dev. Depends on Mean



Arsenic in SRM 1566a: Means and 95% Confidence Intervals



Metrology and the “Bureau”

- Metrology is the science of measurement
- The U.S. national laboratory for measurement science and measurement standards is the **National Institute of Standards and Technology**, formerly the **National Bureau of Standards**
- Interlaboratory studies are central to characterizing measurement systems, and assuring measurement quality; hence NIST/NBS is a center for such investigations.

W.J. Youden and Interlaboratory Comparisons

- Worked with ASTM and AOAC (Association of Official Analytical Chemists)
- Wrote AOAC manual for collaborative tests
- Numerous contributions:
 - Discussions of precision, accuracy, and bias
 - Discussions of design issues
 - Ranking laboratories
 - Outlier test
 - The *Youden Plot*

Youden on Replicates
(“Realistic Estimates of Error,” 1962)

“Repeat measurements cannot reveal the vicissitudes of measurement making unless the operator gives the vicissitudes a chance to occur”

Youden on Controlled Conditions
(“Systematic Errors in Physical Constants,” 1961)

“[Errors calculated under strictly controlled conditions] had no more to do with the real errors than the neatness of the laboratory, or the promptness with which the investigator answered his mail”

Youden Plots

1959: Graphical Diagnosis of Interlaboratory Test Results

1959: Statistical Aspects of the Cement Testing Program

1948: Multiple Factor Experiments in Analytical Chemistry
[precursor?]

The Youden Plot

- Instead of having several laboratories measure the same thing in duplicate, have the labs measure two *similar but different* materials, once each.
- The two measurements determine a point for each lab; these points are displayed in a *Youden Plot*.
- Circular pattern indicates no between-lab variability; points not on 45° line suggest possible problem with measurement or homogeneity.

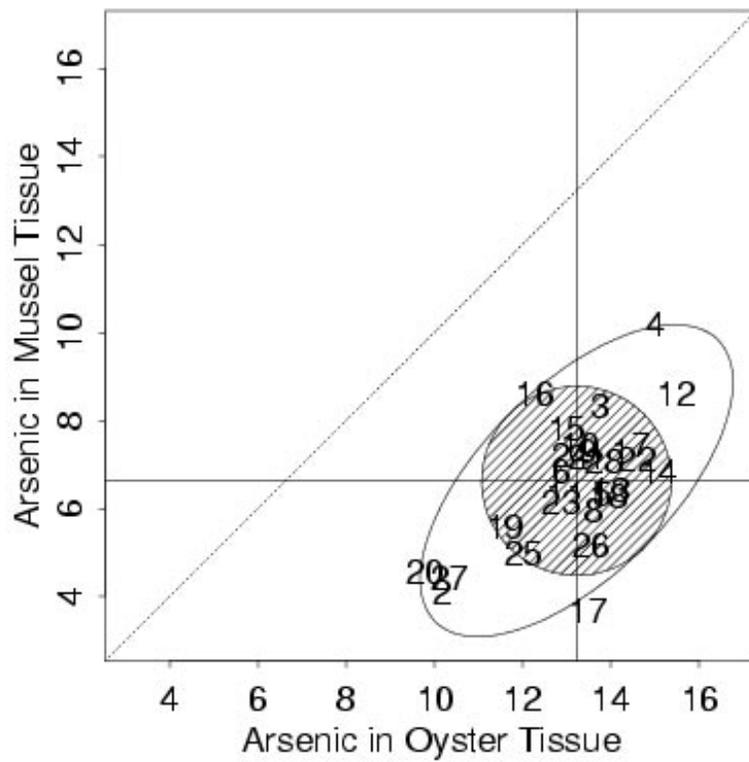
Inference on Youden Plots

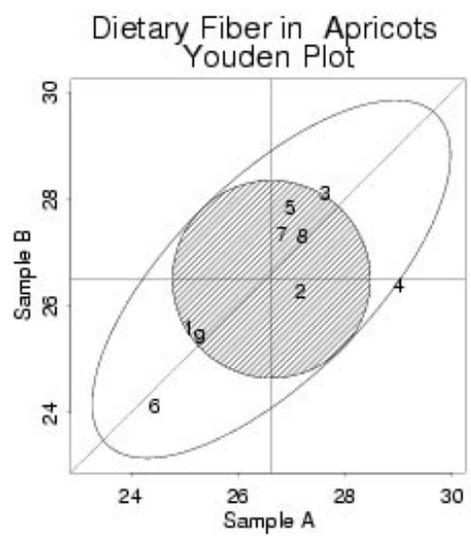
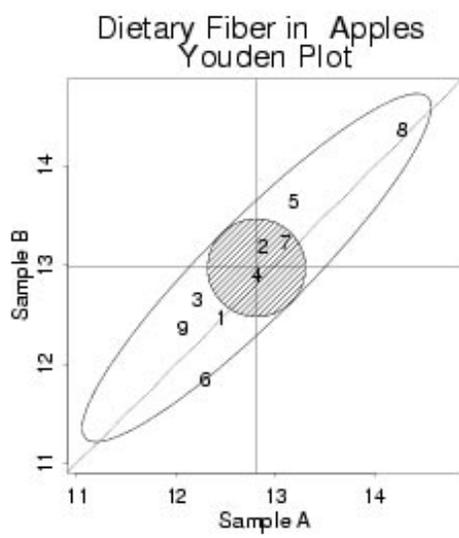
$$\text{Var} \left(\frac{x_{1j} + x_{2j}}{\sqrt{2}} \right) \sim (2\sigma_b^2 + \sigma_e^2) \chi_{p-1}^2 / (p-1)$$

$$\text{Var} \left(\frac{x_{1j} - x_{2j}}{\sqrt{2}} \right) \sim \sigma_e^2 \chi_{p-1}^2 / (p-1)$$

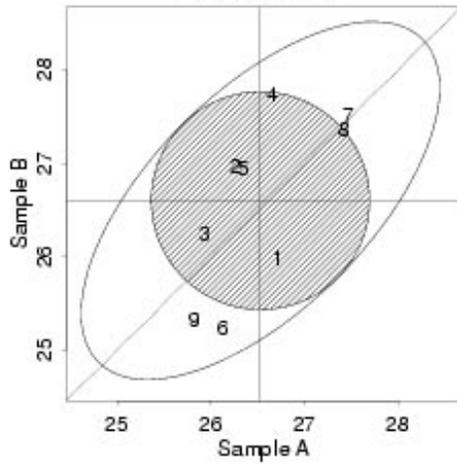
- The ratio of these estimates is the F-statistic for testing $\sigma_b^2 = 0$.
- Confidence and prediction regions are straightforward to calculate.

Arsenic Data From NOAA/NRC Study Ninth Round

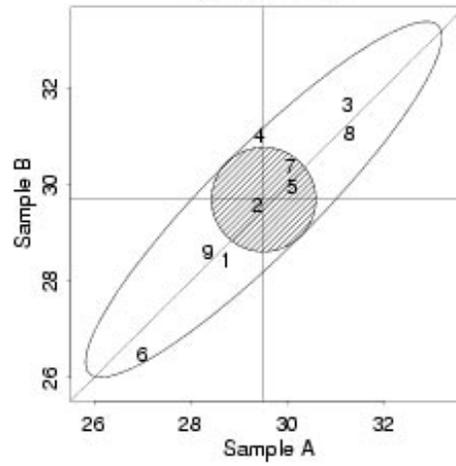


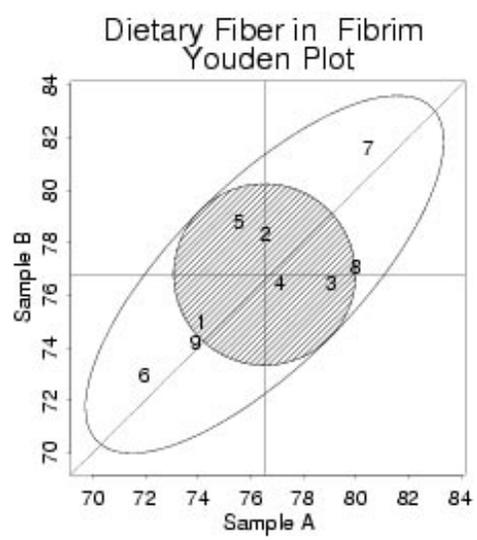
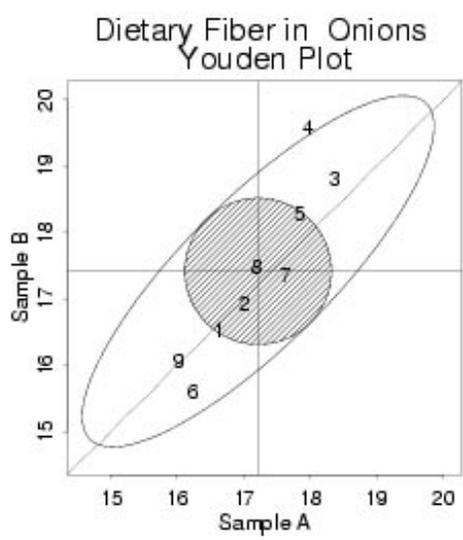


Dietary Fiber in Cabbage
Youden Plot



Dietary Fiber in Carrots
Youden Plot





A Generalized Youden Plot

The objective is to construct a “Youden-like” plot for three equicorrelated responses, hopefully in a way that generalizes to higher dimension. As an example, consider dietary fiber for three foods:

Lab	Apples	Apricots	Cabbage
1	25.32	26.35	28.57
2	26.73	26.62	29.48
3	27.89	26.09	31.46
4	27.70	27.20	30.22
5	27.42	26.67	30.04
6	24.30	25.69	26.76
7	27.11	27.49	30.23
8	27.28	27.41	31.20
9	25.37	25.58	28.49

Model

- Index foods and labs by $i = 1, 2, 3$ and $j = 1, \dots, 9$, respectively. Assume

$$y_{ij} = \mu_i + b_j + e_{ij}$$

where $b_i \sim N(0, \sigma^2)$, $e_{ij} \sim N(0, \sigma_e^2)$.

- Then (x_{1j}, x_{2j}, x_{3j}) is trivariate normal with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma^2 + \sigma_e^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_e^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 + \sigma_e^2 \end{bmatrix}$$

- The eigenvalues of this matrix are σ_e^2 , with multiplicity 2, and $3\sigma^2 + \sigma_e^2$, with multiplicity 1.

Plan

- Consider all planes containing $[1, 1, 1]^T$.
- For each of these, project the data onto this plane.
- Summarize these results in a single plot. It's not immediately obvious how best to do this. One possibility is to show the range of projections for each lab, along with the "average" projection.

Computational Details

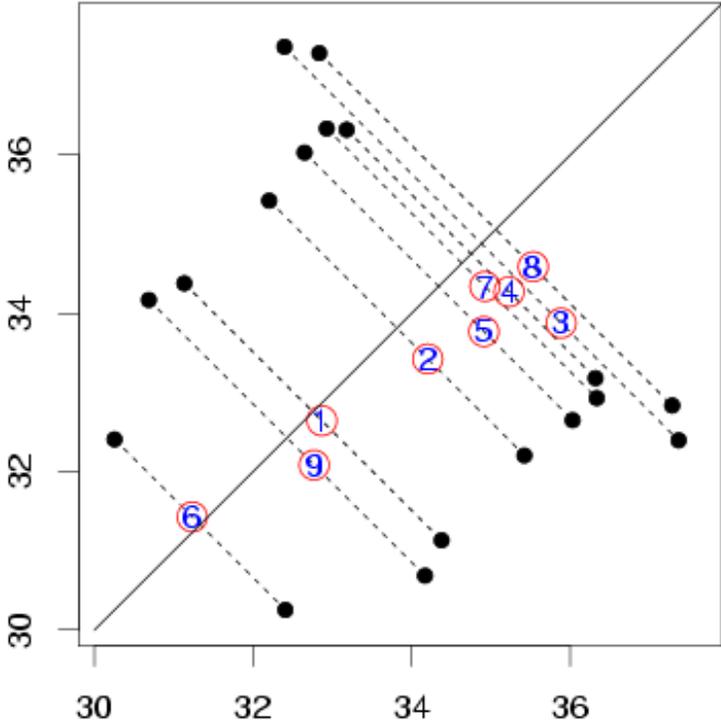
- We need to find all planes containing

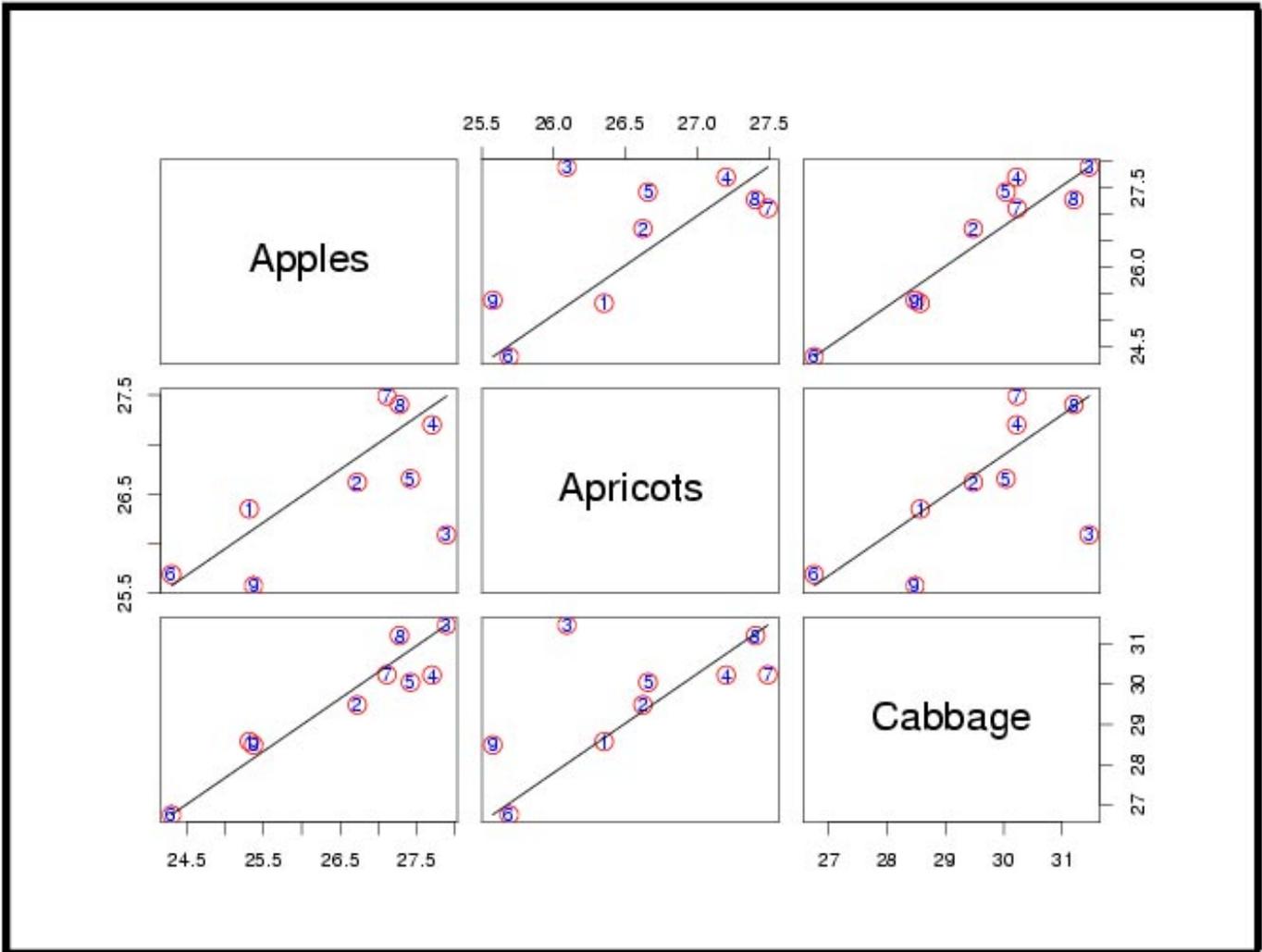
$$J = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$$

- Each such plane is determined by a vector $[\alpha, \beta, \gamma]$ such that $\alpha + \beta + \gamma = 0$ and $\alpha^2 + \beta^2 + \gamma^2 = 1$.
- The solution can be shown to be

$$\begin{aligned} -\sqrt{\frac{2}{3}} &\leq \beta \leq \sqrt{\frac{2}{3}} \\ \alpha &= \frac{-\beta \pm \sqrt{2 - 3\beta^2}}{2} \\ \gamma &= -\alpha - \beta \end{aligned}$$

Generalized Youden Plot: Apples, Apricots, Cabbage





Computation in Higher Dimensions

- In higher dimensions, it becomes more efficient to approximately average over projections by simulation.
- Choose a k -dimensional vector u with *iid* normal components.
- Center and then normalize u ; call the result u' .
- Project the k -dimensional data for each lab onto the plane spanned by J_k and u' .
- Discard any projection which is not within $\pm\pi/2$ of a particular plane containing J_k .
- Average over many such projections.

Ranking Laboratories

- Laboratory ranks are often reported in proficiency studies. (Labs with high or low ranks might be suspect.)
- Youden proposed ranking the labs within each material in a two-way table, and comparing the mean rank with its permutation distribution.
- A Bayesian approach (following Spiegelhalter) is probably preferable.

Ranking Labs: Youden's Approach

Lab.	Food						Sum
	1	2	3	4	5	6	
1	6	6	9	6	6	6	39
2	9	1	6	9	9	9	43
3	3	9	3	1	1	1	18
4	1	2	1	2	2	4	12
5	4	7	2	5	8	5	31
6	2	8	5	4	7	2	28
7	7	5	4	7	5	3	31
8	5	4	8	8	3	8	36
9	8	3	7	3	4	7	32

Rank sums are compared [13, 47] permutation interval.

Hierarchical Model With Noninformative Priors: Two-Way Model

$i = 1, \dots, p$ indexes laboratories

$j = 1, \dots, n_i$ indexes measurements

$k = 1, \dots, m$ indexes materials

$$p(x_{ijk} | \delta_i, \theta_k, \sigma_i^2) = N(\delta_i + \theta_k, \sigma_i^2)$$

$$p(\sigma_i) \propto 1/\sigma_i$$

$$p(\delta_i | \mu, \sigma^2) = N(\mu, \sigma^2)$$

$$p(\theta_k) \propto 1$$

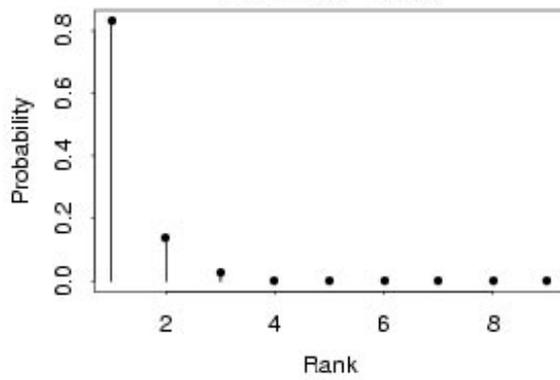
$$p(\sigma) \propto 1$$

BUGS Code for Model Excerpt

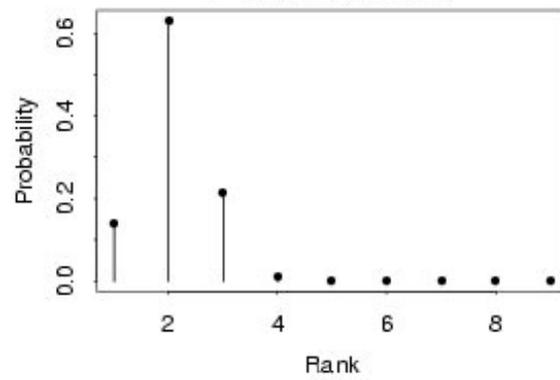
- δ_i is the laboratory “random effect”.
- After each draw from the approximate posterior, determine the ranks of the δ_i by using the functions `step` and `sum`.

```
for (i in 1:LABS){  
  for (j in 1:LABS){  
    greater.than[i,j]<- step(delta[i]-delta[j]);  
  }  
  rank[i]<-sum(greater.than[i,]);  
}
```

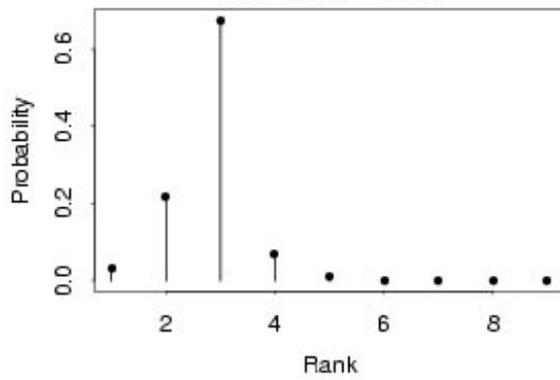
Fiber Data , Lab 6
Posterior Rank



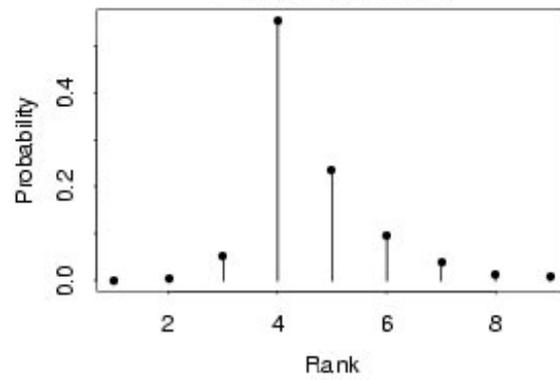
Fiber Data , Lab 9
Posterior Rank

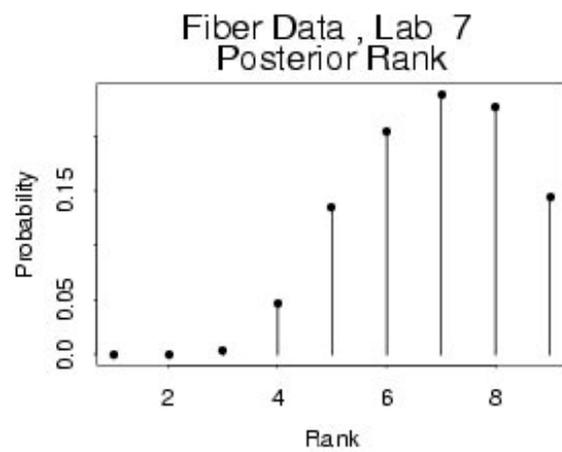
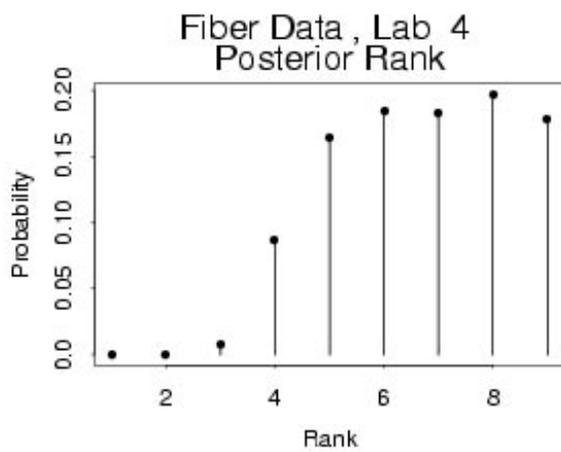
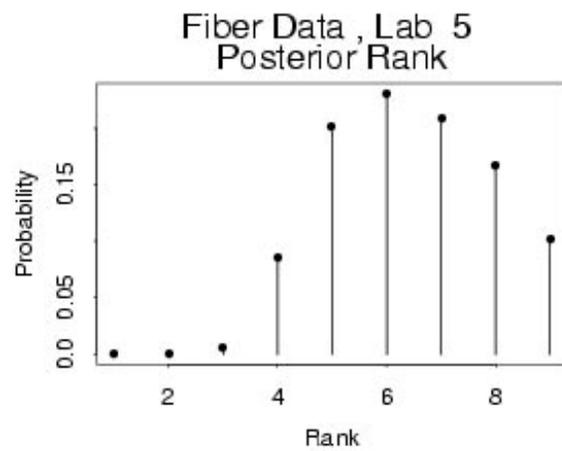
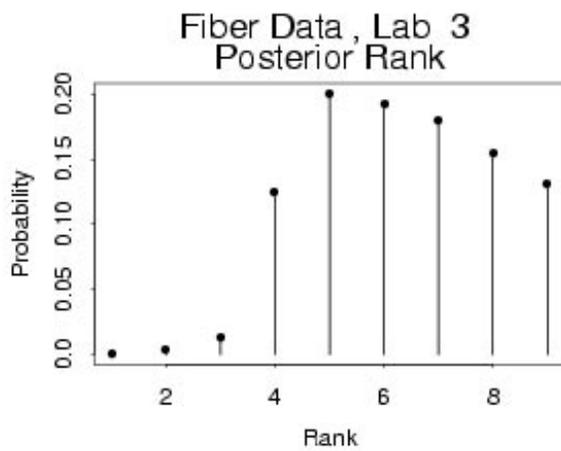


Fiber Data , Lab 1
Posterior Rank

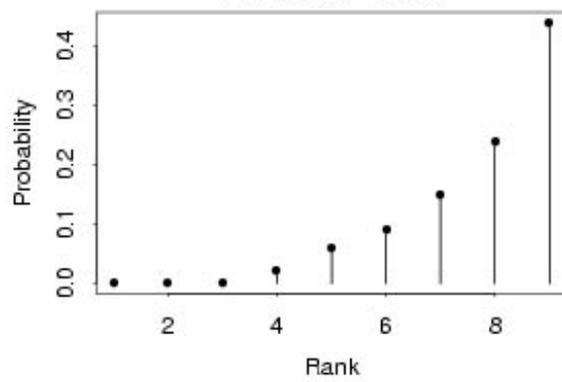


Fiber Data , Lab 2
Posterior Rank

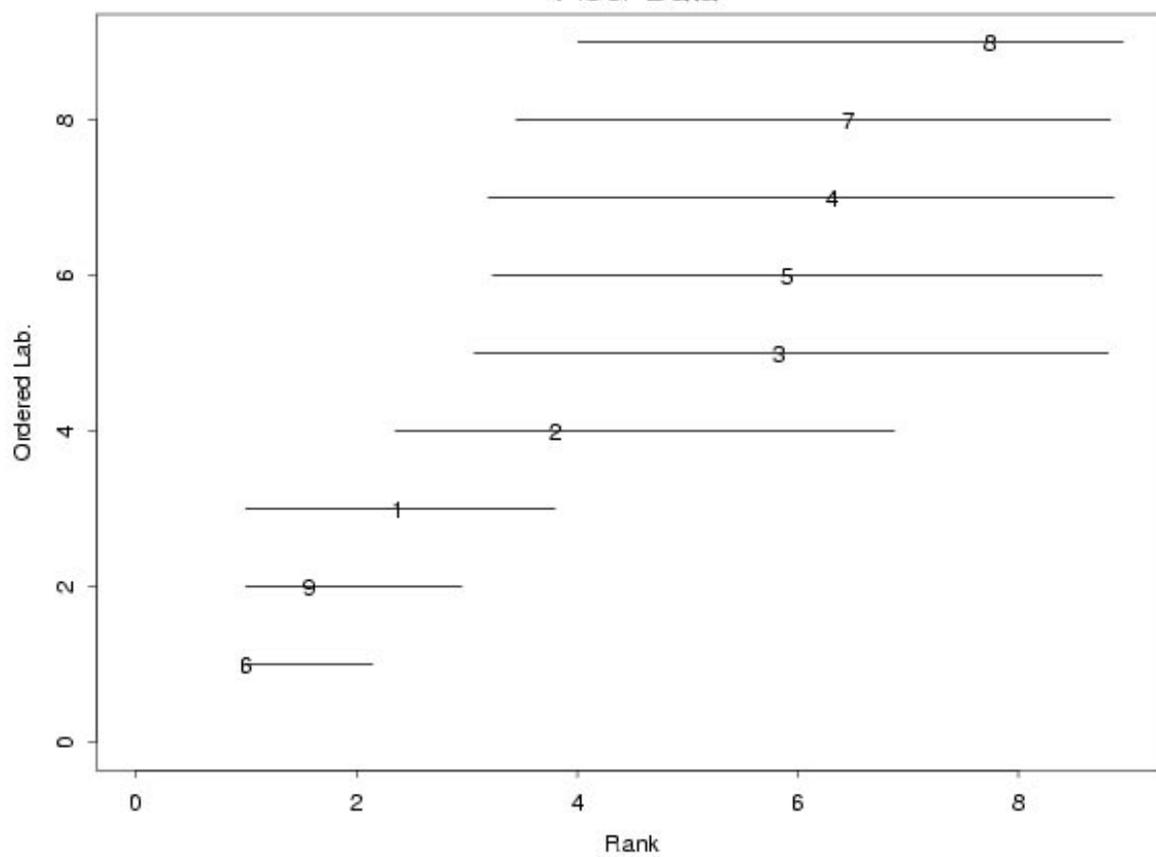




Fiber Data , Lab 8
Posterior Rank



Posterior 95% Intervals for Ranks
Fiber Data



**Interlaboratory Study
Methodology at NBS
John Mandel (1948-)**

- **John Mandel** Also a chemist with statistical training, Mandel's career has been devoted to work in interlaboratory studies, and to understanding measurement as a process. His most influential contributions center on the **Row-Linear Model** for two-way tables in interlaboratory studies, and on the Mandel-Paule approach to single material studies. He has long been an influential member of the American Society for Testing and Materials.

**Single-Material Interlaboratory Model:
One-Way, Unbalanced, Heteroscedastic
Random-Effects ANOVA**

- Laboratory sample means \bar{x}_i distributed independently normal with mean μ and variance $\sigma^2 + \tau_i^2$, where $\tau_i^2 = \sigma_i^2/n_i$.
- Expected mean for i th laboratory is also normal, with mean μ and variance σ^2 .
- Sufficient statistics \bar{x}_i and $t_i^2 = s_i^2/n_i$.

If x_{ij} denotes the j th measurement from the i th lab, then

$$x_{ij} = \mu + b_i + e_{ij},$$

where $b_i \sim N(0, \sigma^2)$ and $e_{ij} = N(0, \sigma_i^2)$; mutually independent.

**One-Way Models in
Interlaboratory Studies:
The Mandel-Paule Estimator
J. of Research of the NBS (1982)**

- For arbitrary positive weights $\{w_i\}_{i=1}^k$,
weighted mean is

$$\tilde{\mu} = \frac{\sum_{i=1}^p w_i \bar{x}_i}{\sum_{i=1}^p w_i}.$$

- *Mandel-Paule* estimate of μ is the weighted mean $\tilde{\mu}$ for which

$$w_i \equiv \frac{1}{\tilde{\sigma}^2 + t_i^2}$$

where $\tilde{\sigma}^2$ is the root (if any) of

$$Q = \sum_{i=1}^p w_i (\bar{x}_i - \tilde{\mu})^2 = p - 1$$

Mandel-Paule Mean for Two Laboratories

- It turns out that the Mandel-Paule estimator of the consensus mean can be found in closed form for $p = 2$ laboratories:
- If

$$\frac{|\bar{x}_{1.} - \bar{x}_{2.}|}{\sqrt{t_1^2 + t_2^2}} \leq p - 1 = 1,$$

then

$$\tilde{\mu} = \frac{\bar{x}_{1.}/t_1^2 + \bar{x}_{2.}/t_2^2}{1/t_1^2 + 1/t_2^2},$$

otherwise

$$\tilde{\mu} = \frac{\bar{x}_{1.} + \bar{x}_{2.}}{2} + \frac{1}{2} \left(\frac{t_1^2 - t_2^2}{\bar{x}_{1.} - \bar{x}_{2.}} \right)$$

The Mandel-Paule Algorithm and ML/REML

Maximum-Likelihood for a linear model

$$Y = X\beta + e,$$

where $e \sim N(0, \Sigma)$ is equivalent to minimizing of $|\Sigma|$, subject to

$$(y - X\hat{\beta})^T \Sigma^{-1} (y - X\hat{\beta}) = n \quad (1)$$

where $\hat{\beta}$ is the GLS estimate of β , and n is the number of observations.

For our one-way model, if the σ_i^2 are replaced by s_i^2 , then (1), an equation in σ^2 alone, is

$$\sum_{i=1}^p w_i (\bar{x}_i - \tilde{\mu})^2 = p.$$

Had REML been used, rather than ML, then the p on the RHS above would be a $p - 1$, *precisely* Mandel and Paule's equation.

Maximum Likelihood (Cochran, 1937)

Let $\omega_i = 1/(\sigma^2 + \tau_i^2)$, $\nu_i = n_i - 1$, and determine $\hat{\sigma}$, $\hat{\tau}_i^2$, and $\hat{\mu}$ to satisfy

$$(A_i) \quad \omega_i - \omega_i^2(\bar{x}_{i.} - \mu)^2 + \nu_i \left(\frac{1}{\tau_i^2} - \frac{t_i^2}{\tau_i^4} \right) = 0$$

$$(B) \quad \boxed{\sum_{i=1}^k \omega_i^2 (\bar{x}_{i.} - \mu)^2 = \sum_{i=1}^k \omega_i}$$

$$(C) \quad \mu = \frac{\sum_{i=1}^k \omega_i \bar{x}_{i.}}{\sum_{i=1}^k \omega_i}$$

Note that (B) may have multiple roots. Cochran (1937) proposed setting $\tau_i^2 = t_i^2$ and solving (B) for σ^2 , then using (C).

The Loglikelihood Function: A Better Parametrization

Define weights by

$$\gamma_i \equiv \frac{\sigma^2}{\sigma^2 + \tau_i^2}$$

The loglikelihood becomes

$$\begin{aligned} 2\ell &= \sum_{i=1}^p n_i \log\left(\frac{\gamma_i}{\sigma^2}\right) \\ &\quad - \sum_{i=1}^p \frac{\gamma_i}{\sigma^2} \left[(\bar{x}_i - \mu)^2 + \frac{\nu_i t_i^2}{1 - \gamma_i} \right] \\ &\quad - \sum_{i=1}^p \nu_i \log(1 - \gamma_i) + K. \end{aligned}$$

Differentiate this with respect to parameters μ, σ^2 and $\gamma_i, i = 1, \dots, p$.

ML Equations

$$\mu = \frac{\sum_{i=1}^p \gamma_i \bar{x}_i}{\sum_i \gamma_i} = \frac{\sum_{i=1}^p \omega_i \bar{x}_i}{\sum_i \omega_i}$$

$$\sigma^2 = \frac{\sum_{i=1}^p \gamma_i \left[(\bar{x}_i - \mu)^2 + \frac{\nu_i t_i^2}{1 - \gamma_i} \right]}{\sum_{i=1}^p n_i}$$

$$\begin{aligned} & \gamma_i^3 - (a_i + 2)\gamma_i^2 + \\ & [(n_i + 1)a_i + (n_i - 1)b_i + 1] \gamma_i \\ & - n_i a_i = 0 \end{aligned}$$

where

$$a_i \equiv \frac{\sigma^2}{(\bar{x}_i - \mu)^2}$$

and

$$b_i \equiv \frac{t_i^2}{(\bar{x}_i - \mu)^2}.$$

Result #1:
Monotone Convergence to Stationary Points of the Likelihood

- For any starting values μ_0, σ_0^2 , maximize the likelihood over the weights by solving the cubics. (If there are multiple real roots, choose the one which causes the biggest increase in the likelihood.)
- Let

$$\sigma_1^2 = \frac{\sum_{i=1}^p \gamma_i \left[(\bar{x}_i - \mu)^2 + \frac{\nu_i t_i^2}{1 - \gamma_i} \right]}{\sum_{i=1}^p n_i}$$
$$\mu_1 = \frac{\sum_{i=1}^p \gamma_i \bar{x}_i}{\sum_{i=1}^p \gamma_i}$$

solve for new weights, and iterate. This iteration always monotonically increases the likelihood.

Result #2:
Location of Stationary Values of the Likelihood

- At a stationary point of the likelihood,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^p \gamma_i^2 (\bar{x}_{i.} - \mu)^2}{\sum_{i=1}^p \gamma_i}$$

hence

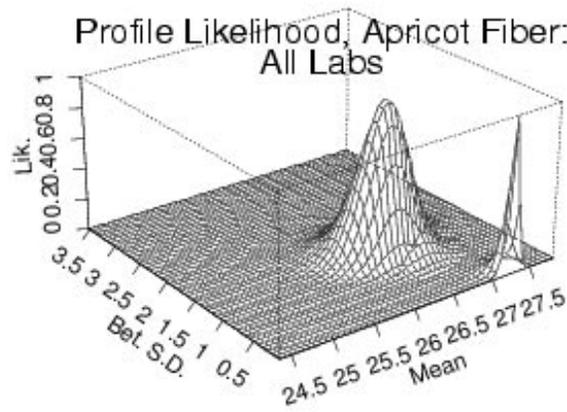
- *All* of the stationary points of the likelihood $\hat{\mu}$ and $\hat{\sigma}$ are within the rectangle in the (μ, σ) plane given by

$$\min_i(\bar{x}_{i.}) \leq \tilde{\mu} \leq \max_i(\bar{x}_{i.})$$

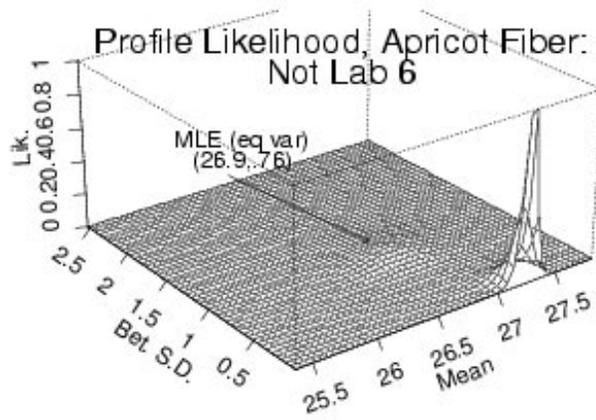
and

$$0 \leq \tilde{\sigma} \leq \max_i(\bar{x}_{i.}) - \min_i(\bar{x}_{i.}).$$

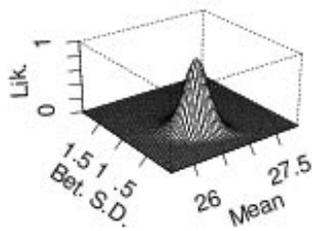
Profile Likelihood, Apricot Fiber:
All Labs



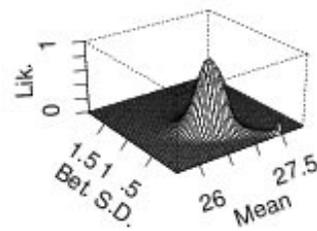
Profile Likelihood, Apricot Fiber:
Not Lab 6



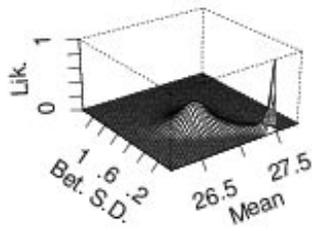
Prof. Lik., Cabbage Fiber:
All Labs



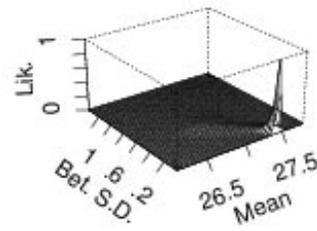
Prof. Lik., Cabbage Fiber:
Not Lab 6



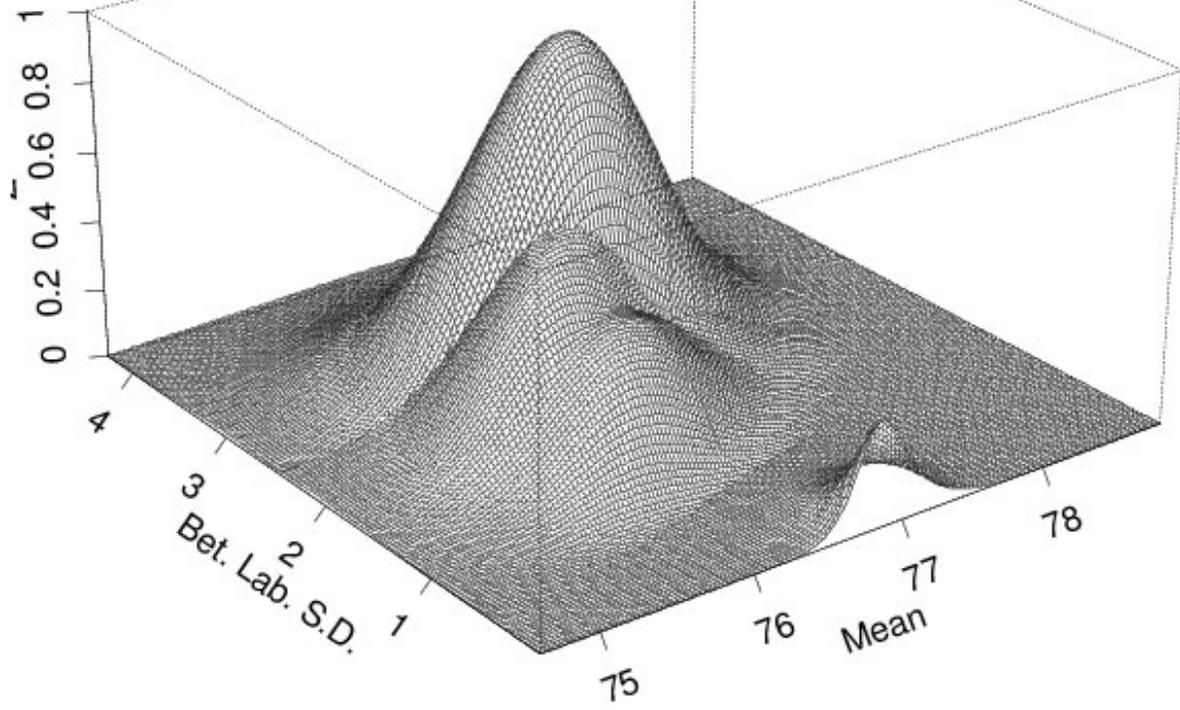
Prof. Lik., Cabbage Fiber:
Not Labs 6, 9



Prof. Lik., Cabbage Fiber:
Not Labs 6, 9, 1



Likelihood for Dietary Fiber in FIBRIM



Result #3:

Location of the Roots of Cubic Equations for Weights (γ_i)

- Each cubic likelihood equation has one or three roots $\gamma_i \in [0, 1]$.
- The “best” necessary condition for three roots is that

$$(\bar{x}_i - \mu)^2 \geq \max(\sigma^2/q_i, t_i^2/h_i),$$

where

$$\begin{aligned} q_i &= -2 - 6\sqrt{n_i} \sin \left\{ \frac{1}{3} \left[\sin^{-1} \left(\sqrt{\frac{n_i - 1}{n_i}} \right) - \frac{\pi}{2} \right] \right\} \\ &= \frac{8}{27n_i} + O(n_i^{-2}) \end{aligned}$$

and

$$h_i = \frac{(1 - q_i)^3}{27(n_i - 1)} = \frac{1}{27n_i} + O(n_i^{-2}).$$

Hierarchical Model With Noninformative Priors

$i = 1, \dots, p$ indexes laboratories

$j = 1, \dots, n_i$ indexes measurements

$$p(x_{ij}|\delta_i, \sigma_i^2) = \text{N}(\delta_i, \sigma_i^2)$$

$$p(\sigma_i) \propto 1/\sigma_i$$

$$p(\delta_i|\mu, \sigma^2) = \text{N}(\mu, \sigma^2)$$

$$p(\mu) \propto 1$$

$$p(\sigma) \propto 1$$

Posterior given $\sigma = 0$, $p \geq 1$

Given $\sigma = 0$, then the posterior distribution of the consensus mean μ is proportional to a product of scaled t -densities:

$$p(\mu|\{x_{ij}\}|\sigma = 0) \propto \prod_{i=1}^p \frac{1}{t_i} T'_{n_i-1} \left(\frac{\bar{x}_{i.} - \mu}{t_i} \right)$$

The General Case: $\sigma \geq 0$

In general, $p(\mu|\sigma, \{x_{ij}\})$ is proportional to a *product* of the distributions of the random variables

$$U_i = \bar{x}_i + \frac{\sqrt{n_i}}{s_i} T_{n_i-1} + \sigma Z,$$

where T_{n_i-1} is a t -distributed random variable with $n_i - 1$ degrees of freedom, Z is distributed $N(0, 1)$, and T_{n_i-1} and Z are independent.

A Useful Probability Density

Let T_ν and Z denote independent Student- t and standard normal random variables, and assume that $\psi \geq 0$ and $\nu > 0$. Then

$$U = T_\nu + Z\sqrt{\frac{\psi}{2}}$$

has density

$$f_\nu(u; \psi) \equiv \frac{1}{\Gamma_{\nu/2}\sqrt{\pi}} \int_0^\infty \frac{y^{(\nu+1)/2-1} e^{-y\left[1+\frac{u^2}{\psi y+\nu}\right]}}{\sqrt{\psi y+\nu}} dy.$$

Posterior of (μ, σ)

- Assume $\delta_i \sim N(\mu, \sigma^2)$, $\sigma \sim p(\sigma)$,
 $p(\mu) \propto 1$, $p(\sigma_i) \propto 1/\sigma_i$.
- Then the posterior of (μ, σ) is

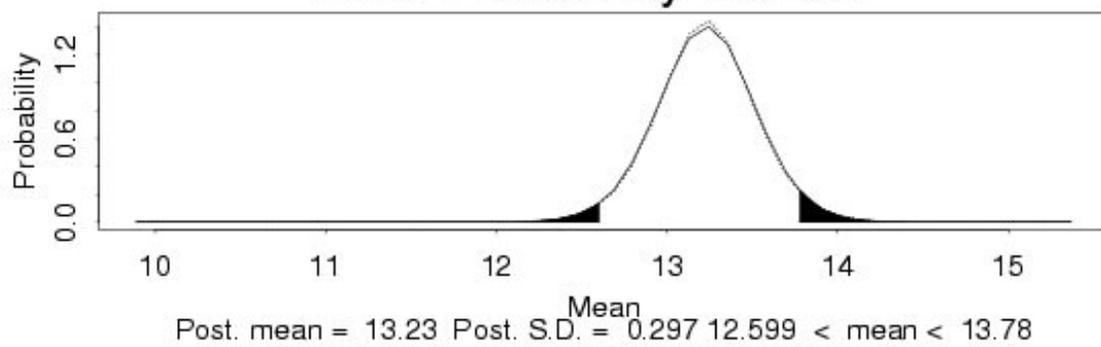
$$p(\mu, \sigma | \{x_{ij}\}) \propto p(\sigma) \prod_{i=1}^p \frac{1}{t_i} f_{n_i-1} \left[\frac{\bar{x}_{i.} - \mu}{t_i}; \frac{2\sigma^2}{t_i^2} \right].$$

- The posterior of μ given $\sigma = 0$ is a product of scaled t -densities centered at the $\bar{x}_{i.}$, since

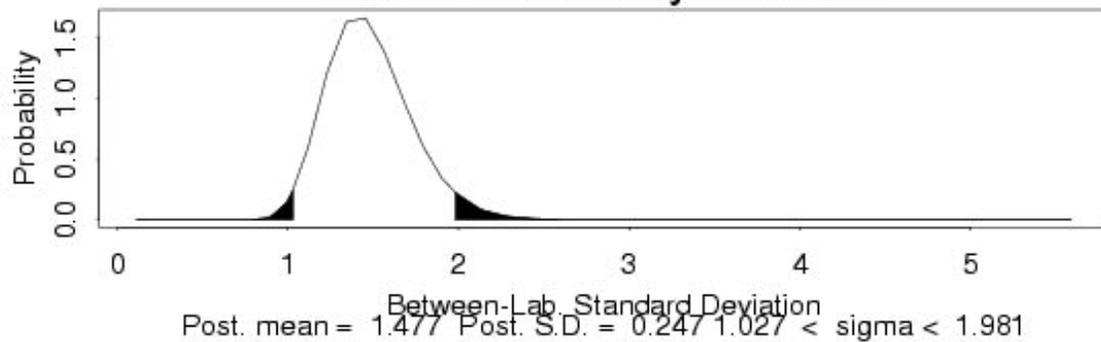
$$\frac{1}{t_i} f_{n_i-1} \left[\frac{\bar{x}_{i.} - \mu}{t_i}; 0 \right] = \frac{1}{t_i} T'_{n_i-1} \left(\frac{\bar{x}_{i.} - \mu}{t_i} \right).$$

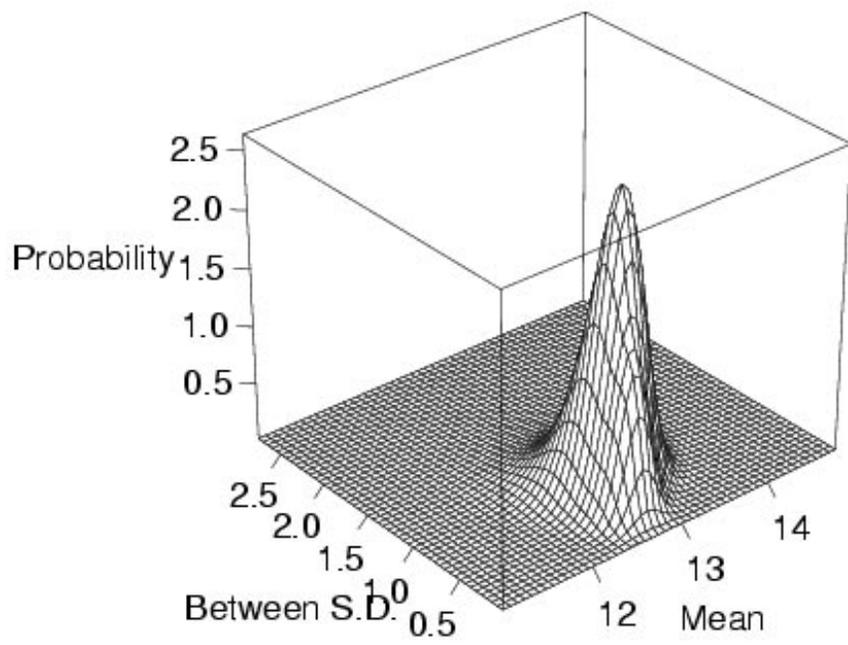
- We will take $p(\sigma) = 1$, though an arbitrary proper prior does not introduce additional difficulties.

Marginal Posterior of Mean With 95% Probability Interval



Marginal Posterior of Between-Lab. S.D. With 95% Probability Interval





**Small Simulation Comparing
Bayesian and Frequentist Intervals**

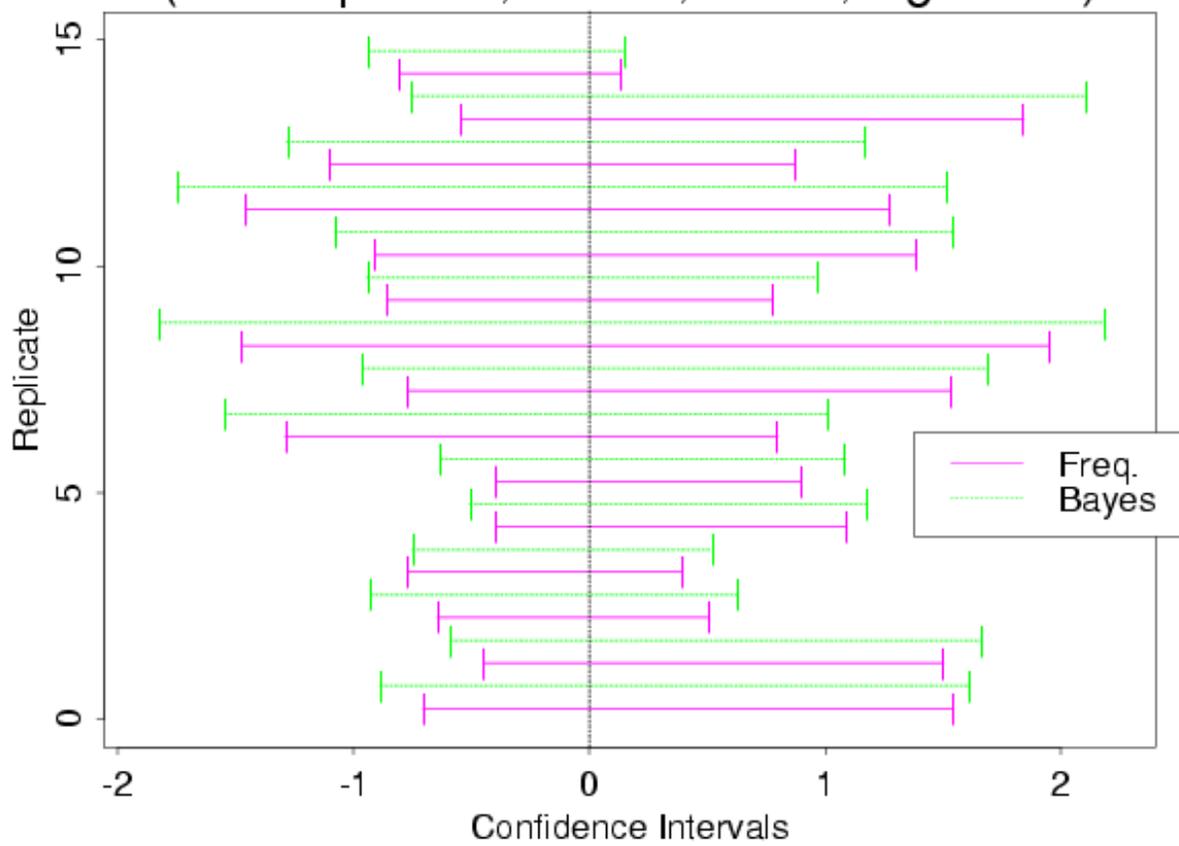
$$\mu = 0$$

$$\sigma_i = \sigma_e$$

$$\sigma^2 + \sigma_e^2 = 1$$

$$\rho = \sigma^2 / (\sigma_e^2 + \sigma^2) = 1/2$$

Simulation Comparing Confidence Intervals (5 Groups of 5, $\rho=.5$, $\mu=0$, $\sigma=1$)



Two-Way Tables

- The typical data structure for an interlaboratory study is a two-way table, although sometimes (as above) the data are analyzed one material at a time.
- One way to model such data is a two-way ANOVA with interaction:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where

- $\alpha_i, i = 1, \dots, p$ is the *laboratory effect* (perhaps random)
- $\beta_j, j = 1, \dots, m$ is the *material effect* (fixed)
- γ_{ij} is the lab/material interaction
- $e_{ijk}, k = 1, \dots, n_{ij}$ is the measurement error, with variance which probably depends on material.

Mandel's Approach to Two-Way Tables

- Typically, one sees unequal error variances for different materials, and often nonadditivity as well.
- Transforming the data can help, but Mandel argues that this is not appropriate since there are multiple variances in the model.
- Mandel's approach consists of
 1. Estimating the within variance separately for each material, and then reducing the data to cell means.
 2. Estimating the row effects, column effects, and interaction.
 3. Regressing the estimated interaction against the column (material) effects. This results in a decomposition of the interaction into a part due to slopes among labs, and a residual.

Calculations for the Row-Linear Model

- Error variances:

$$s_j^2 = \frac{\sum_{i=1}^p \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{.j})^2}{\sum_{i=1}^p (n_{ij} - 1)}$$

- Effects:

$$\begin{aligned}\hat{\mu} &= \bar{y}_{...} \\ \hat{\alpha}_i &= \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} \\ \hat{\gamma}_{ij} &= \bar{y}_{ij.} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j\end{aligned}$$

- Row-linear model for interaction:

$$\hat{\gamma}_{ij} = b_i \hat{\beta}_j + h_{ij}$$

where b_i is the least-squares slope for the i th lab., and h_{ij} is the part of the interaction not explained by the linear regression.

Laboratory Linear Regressions for the Fiber Data

- If we do these linear regressions for the fiber data, we find some very significant slopes. But also some insignificant ones.
- The significant slopes are strongly influenced by the Fibrin data.

Lab.	\hat{b}_i	$s_{\hat{b}_i}$	$\hat{b}_i/s_{\hat{b}_i}$	P-Value
1	-0.0256	0.0075	-3.4129	0.0270
2	0.0140	0.0044	3.1827	0.0334
3	0.0123	0.0207	0.5940	0.5845
4	-0.0095	0.0115	-0.8274	0.4545
5	-0.0005	0.0058	-0.0814	0.9391
6	-0.0504	0.0137	-3.6760	0.0213
7	0.0686	0.0064	10.7495	0.0004
8	0.0183	0.0120	1.5239	0.2022
9	-0.0272	0.0041	-6.5961	0.0027

The Row-Linear ANOVA Table

We can write Mandel's model as:

$$\bar{y}_{ij.} = \bar{y}_{i..} + (b_i + 1)(\bar{y}_{.j.} - \bar{y}_{...}) + h_{ij}$$

Some refer to this as Mandel's 'bundle-of-lines'.

The ANOVA table is

Rows	$p - 1$	$m \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$
Columns	$m - 1$	$p \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$
Interaction	$(p - 1)(m - 1)$	$\sum_{ij} (y_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2$
Slopes	$p - 1$	$\sum_{ij} b_i^2 (\bar{y}_{.j.} - \bar{y}_{...})^2$
Remainder	$(p - 1)(m - 2)$	$\sum_{ij} h_{ij}^2$

Row-Linear ANOVA Table for Dietary Fiber Data

Source	SS	df	F-Ratio
Labs.	64.48	8	23.08
Foods	12447871.91	5	7127822.96
Interaction	36.04	40	2.58
Slopes	24.86	8	8.90
Resid.	11.18	32	

P-Value for Slopes: 2.5×10^{-6}

Summary

- Interlaboratory studies are important in many fields.
- W.J. Youden and John Mandel of the National Bureau of Standards have left an important mark on the simple methods in common use today, through the Youden plot, the Row-Linear Model, and other ideas.
- These methods have been reviewed, and some extensions presented.
- There remains considerable opportunity for new methodology, more realistic and computationally intensive.