STATISTICAL FITTING OF ADSORPTION ISOTHERMS

Andrew R. Barron, Lidija Jakobek Barron, Mirta Benšić, Petra Matić

A. R. Barron — Department of Statistics and Data Science, Yale University

Lidija Jakobek Barron, Mirta Benšić, Petra Matić— University of Osijek

Zagreb, Jun 2019.

This work was supported by the Croatian Science Foundation research grants IP-2016-06-6545 and IP-2016-06-6777

Experiment

- The aim of the experiment — to study the influence of dietary fiber on polyphenol bioaccessibility
- Bioaccessibility — the amount of polyphenols that are actually accessible for absorption in the digestive tract
- Dietary fibers are resistant to digestion and absorption in the human small intestine.
- They have the potential to interact with polyphenols in the digestive tract, to „trap" and „carry" them unchanged to the lower parts of the digestive tract and affect their bioaccessibility.
- The influence of dietary fibers on polyphenol bioaccessibility can be studied by studying the adsorption process between fibers and polyphenols.
- Adsorption proces in a model: **polyphenol** - β-**glucan** (a dietary fiber found in cereals — barley and oats).

Adsorption isotherms

- Particles (an initial concentration) are introduced into a medium (liquid or gaseous at a particular temperature) with an adsorbent present.
- After sufficient time to allow an equilibrium of adsorption to be reached, concentration not adsorbed is observed.
- From the initial concentration and the concentration not adsorbed, the amount adsorbed is calculated.
- A key aim is to know how the amount adsorbed is related to the concentration not adsorbed in the medium for a specific temperature.

Notations

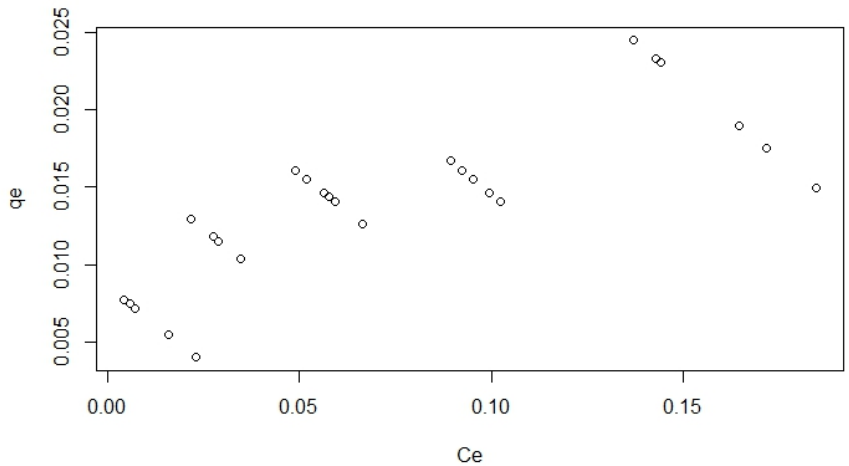
- An initial concentration M (mg/l) is set to one of various specified experimental values.
- After equilibrium is reached the remaining concentration C_e in mg/l is measured. It depends on M.
- From C_e, the amount adsorbed q_e (in mg per g of adsorbent) is determined using the principle of conservation of mass by the material balance equation

$$q_e = (M - C_e)/\gamma,$$

  where γ is the amount of adsorbent in g/l.
- $q_e \longleftrightarrow C_e$???

Data

- Observed $(C_{e,i}, q_{e,i})$, $i = 1, \ldots, n$ (n — number of trials)
- Plot reveals the experimental structure — groups of collinear points (one group for each initial condition M) with slope $-1/\gamma$



Equilibrium equations

Equilibrium equations have been most frequently employed to interpret adsorption data.

$$q_e - \text{amount adsorbed}$$

$$C_e - \text{concentration not adsorbed}$$

- Freundlich model:

$$q_e = K(C_e)^a$$

- Langmuir model:

$$q_e = Q\frac{C_e}{K_L + C_e}$$

- Dubinin-Radushkevich model:

$$q_e = Q_s e^{-a(\log \frac{C_s}{C_e})^2}$$

(K, a), (K_L,Q) and (Q_s,C_s) — constants for a given adsorbate and adsorbent at a particular temperature.

Estimation - standard procedure

- Data $(C_{e,i}, q_{e,i})$, $i = 1, \ldots, n$
- Model: $q_e = h(C_e, \boldsymbol{a})$, $\boldsymbol{a}$ unknown vector parameter

Nonlinear least squares — parameter value $\hat{\boldsymbol{a}}$ minimizes the (weighted) sum of squares of error

$$\sum_{i=1}^{n} w_i (q_{e,i} - h(C_{e,i}, \boldsymbol{a}))^2.$$

Estimation - standard procedure - problem

- Structural relationships:

$$q_e = h(C_e, \boldsymbol{a})$$

$$q_e = (M - C_e)/\gamma$$

- Measurements:

$$C_{e,i} = C_e + \varepsilon_i$$

$$q_{e,i} = (M - C_{e,i})/\gamma = (M - C_e - \varepsilon_i)/\gamma = q_e + \nu_i$$

- From the structural form we have

$$q_{e,i} = h(C_e, \boldsymbol{a}) + \nu_i$$

$$q_{e,i} = h(C_{e,i} - \varepsilon_i, \boldsymbol{a}) + \nu_i$$

- joint variability of $C_{e,i}$ and $q_{e,i}$, $E(\nu_i C_{e,i}) \neq 0$???
- endogenity - problem with inconsistency even in the linear LS

Proper statistical fitting

- Experiments vary a precisely controlled choice of an initial mass M: $m_1, \ldots, m_n$
- C_e is related to M
- Rearrange the models to express $q_e = f(M, \boldsymbol{a})$ as a function of the original input M by solving the material balance equation

$$q_e - (M - h^{-1}(q_e))/\gamma = 0$$

- Now we have a classical nonlinear regression but with M as a predictor.

$$q_{e,i} = f(m_i, \boldsymbol{a}) + \epsilon_i$$

Proper statistical fitting - problem

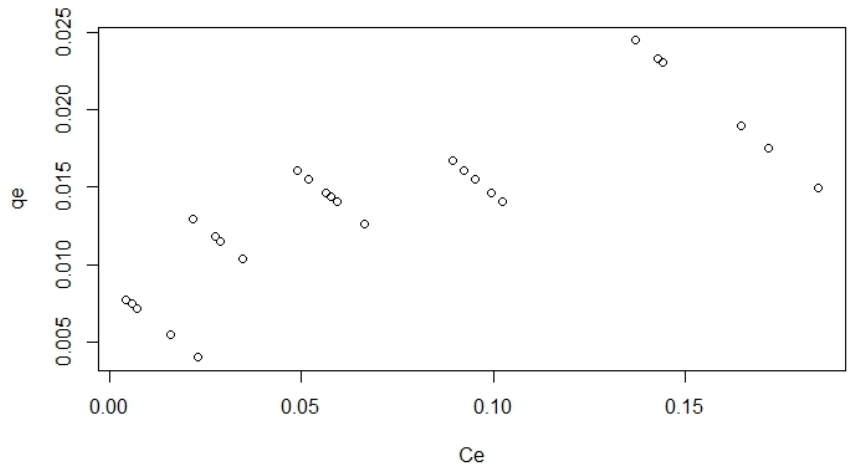
- For Langmuir model, f can be derived explicitly
- Otherwise, f is implicitly given by the equation

$$q_e - (M - h^{-1}(q_e))/\gamma = 0$$

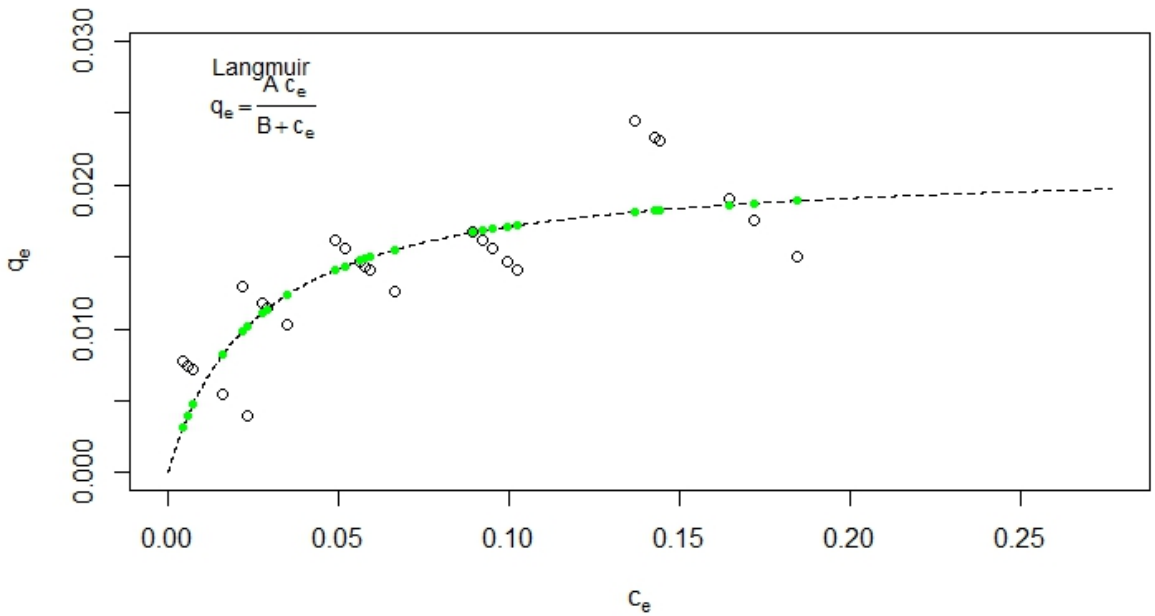
- Computation in standard statistical packages — problem if we don't have an explicit expression for f (estimations, se, CI)

Experiment and data

- Data: initial mass M at 5 levels: 25,50,75,100,150
- At each mass level, 6 experiments
- Langmuir model



Standard procedure LS



The fitted (C_e, q_e) does not correspond to the experimentally given initial mass.

Improved procedure LS

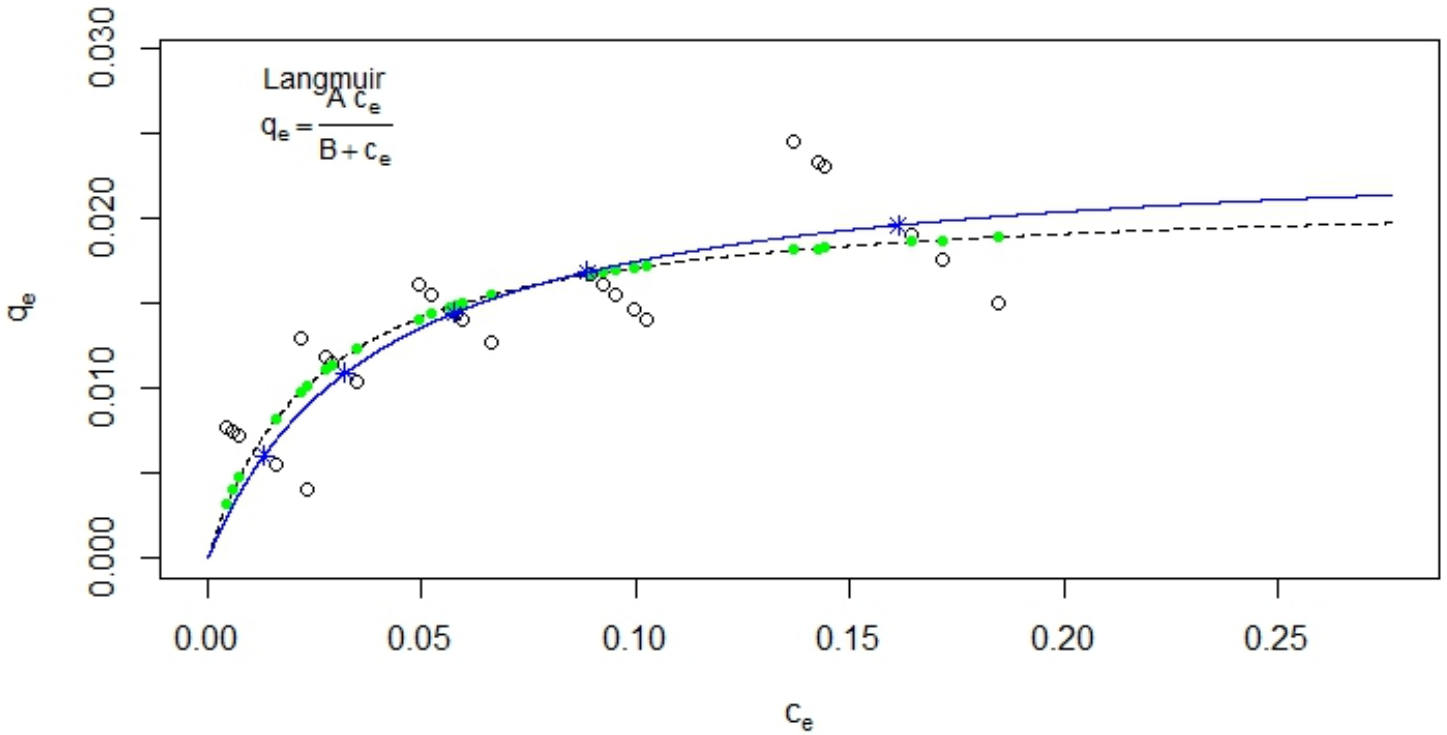
- Define $f(M, \boldsymbol{a})$ as a solution of the equation

$$q_e - (M - h^{-1}(q_e, \boldsymbol{a}))/\gamma = 0$$

- Minimize

$$S(\boldsymbol{a}) = \sum_{i=1}^{n} w_i (q_{e,i} - f(m_i, \boldsymbol{a}))^2.$$

Improved procedure LS



Improved procedure LS

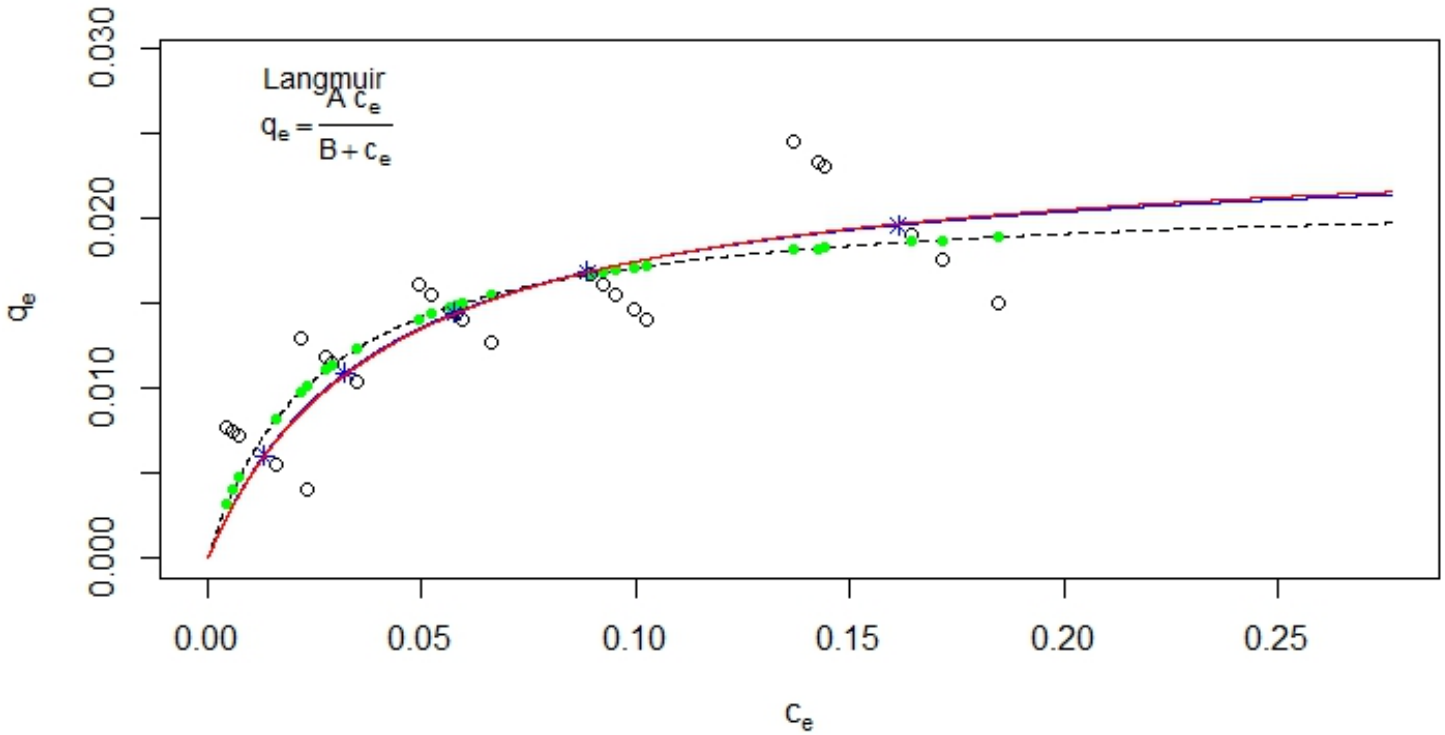
- The estimator is asymptotically normal.
- We computed the asymptotic variance and bias.
- Variance:

$$\left[\sum_{i=1}^{n} w_i \nabla f(M_i, \boldsymbol{a}) \nabla f(M_i, \boldsymbol{a})^\tau\right]^{-1} =$$

$$= \left[\sum_{i=1}^{n} w_i \frac{1}{(1 + \frac{1}{\gamma}\frac{\partial h}{\partial Ce}(Ce_i))^2} \nabla h(Ce_i, \boldsymbol{a}) \nabla h(Ce_i, \boldsymbol{a})^\tau\right]^{-1}$$

- Bias correction — Huge expression with (mostly) small impact.

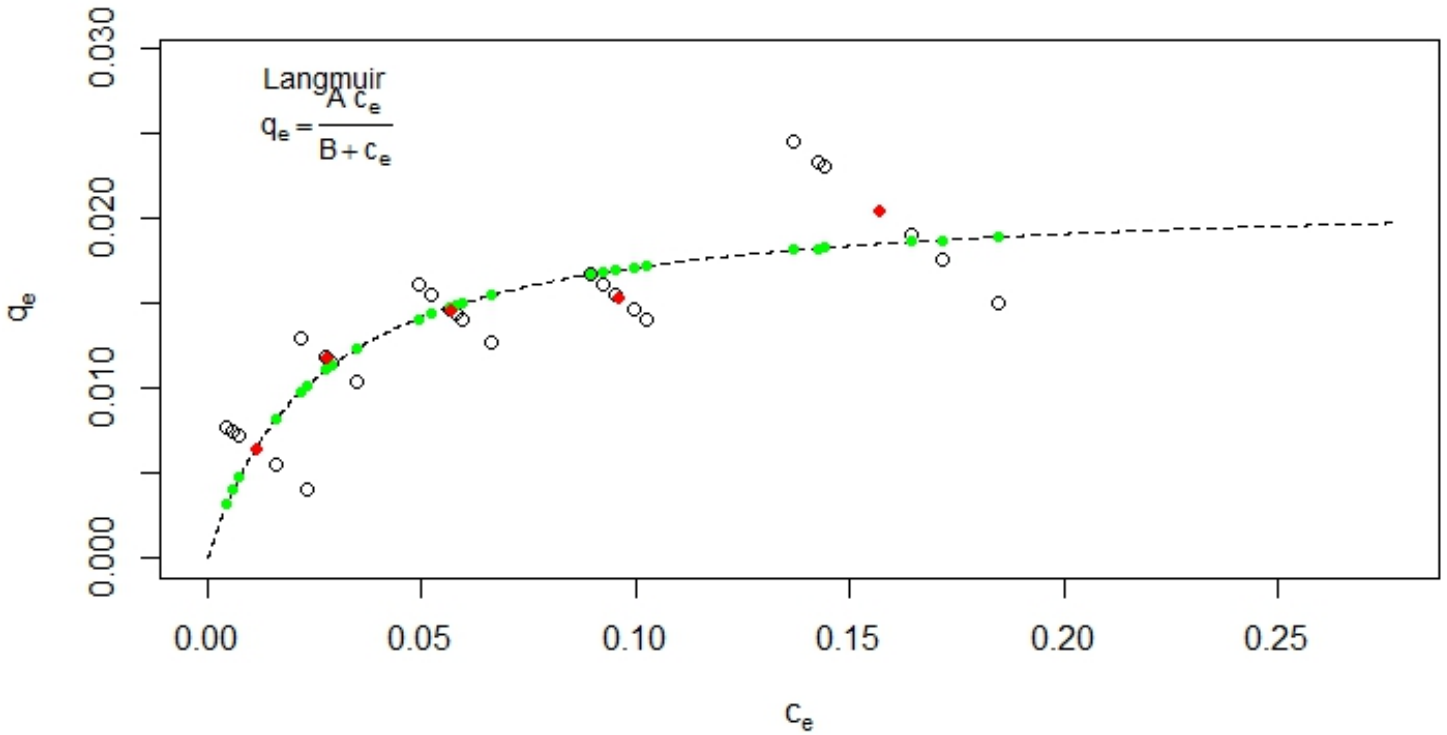
Bias corrected improved procedure LS



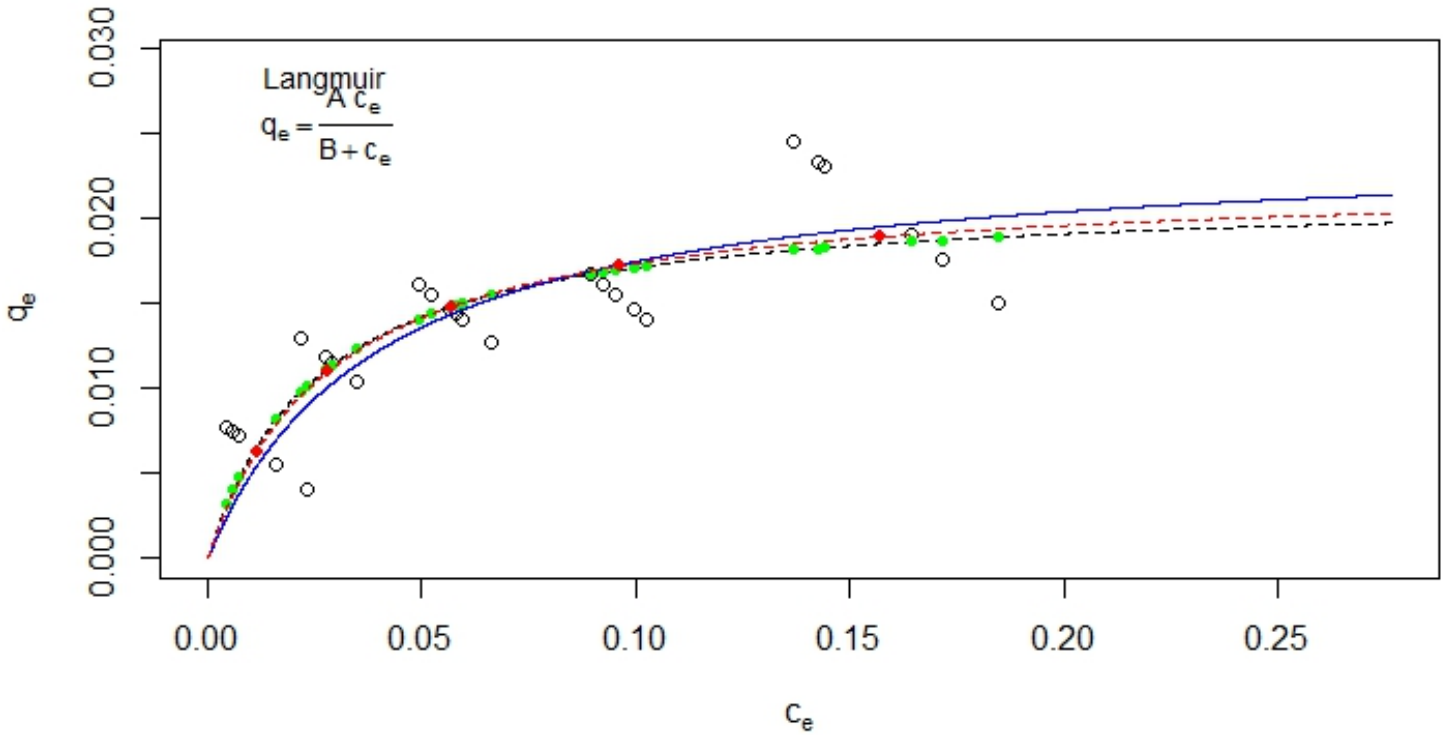
Fit from means

- Repeated experiments for the chosen levels of initial mass M.
- Use average values for C_e and q_e at each level and standard LS procedure.
- Slightly improves the naive fit.
- Asymptotically OK.

Fit from means



Fit from means



Results for polyphenol - β-glucan data

estimator	Standard	From Means	Improved
sd residuals	0.00295	0.00224	0.00211

estimator	Standard	From Means	Improved	Im. Bias Corr.
A	0.0216	0.0224	0.0245	0.0248
		(0.00226)	(0.00232)	(0.00232)
B	0.0261	0.0293	0.0401	0.0423
		(0.0118)	(0.0125)	(0.0125)

Small sample simulation

Langmuir model

$$A = 0.02,\ B = 0.03$$

$$M = (25, 50, 75, 100, 150)$$

mass equation: $580 * C_e + 2900 * q_e = M$
C_e computed from M:

$$580 * C_e + 2900 * \frac{AC_e}{B + C_e} = M$$

$$C_e = (0.01294334, 0.03347204, 0.06195567, 0.09621245, 0.17338687)$$

Data — 6 data for each M level, C_e plus normal errors, $sd = 0.0081$
10000 replications

Results

$$A = 0.02$$

estimator	Standard A	From Means A	Improved A	Im. Bias Corr. A
mean	0.019369	0.020241	0.020212	0.020069
se	0.001496	0.001555	0.001384	0.001286

$$B = 0.03$$

estimator	Standard B	From Means B	Improved B	Im. Bias Corr. B
mean	0.028098	0.031941	0.031481	0.030538
se	0.008872	0.009385	0.008288	0.007537

Confidence segments for (C_e, q_e) for improved procedure

- Based on the estimator's asymptotic normality and computed standard errors.
- For fixed M, confidence intervals can be calculated for C_e (let say I_{C_e}) and q_e (I_{q_e}).
- Each pair (C_e, q_e) should satisfy the mass equation.
- Using I_{C_e} and I_{q_e}, a confidence segment at the mass equation line can be detected for each M.

Improved procedure LS with confidence bands