Longitudinal concordance correlation function and applications

Thiago de Paula Oliveira

thiago.paula.oliveira@insight-centre.org
https://prof-thiagooliveira.github.io/AcademicProfile/

May 1, 2019

Contents

1 Concordance correlation coefficient

2 Longitudinal Concordance Correlation Function

3 Motivation: The papaya's hue colour analysis
- Objective
- Experimental summary
- Definition of longitudinal concordance correlation function
- Results

4 The `lcc` package

5 References

Concordance correlation coefficient

- Concordance correlation coefficient (CCC) to evaluate reproducibility (Lin, 1989):
  - Validation process: evaluate whether the new assay can reproduce the results based on a traditional gold-standard assay;

Definition 1. Concordance correlation coefficient (Lin, 1989)

Let us assume that pairs of samples (Y_{i1}, Y_{i2}), $i = 1, 2, \ldots, n$, are independent selected from a bivariate population with means μ_1 and μ_2 and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Concordance correlation coefficient

Definition 1. Concordance correlation coefficient (Lin, 1989)

Thus, the CCC for two methods based on variances, covariance and means is defined as follows:

$$\rho_c = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \rho C_b, \quad -1 \leq \rho_c \leq 1 \tag{1}$$

Pearson correlation coefficient ρ

$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_1^2 \sigma_2^2}}$ is the Pearson correlation coefficient (measures how far each observation deviated from the best-fit line (precision measure))

Concordance correlation coefficient

Definition 1. Concordance correlation coefficient (Lin, 1989)

Thus, the CCC for two methods based on variances, covariance and means is defined as follows:

$$\rho_c = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \rho C_b, \quad -1 \leq \rho_c \leq 1 \tag{1}$$

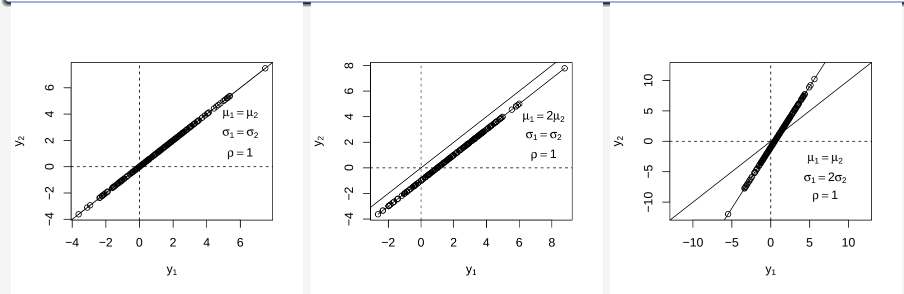
Pearson correlation coefficient ρ

$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$ is the Pearson correlation coefficient (measures how far each observation deviated from the best-fit line (precision measure))

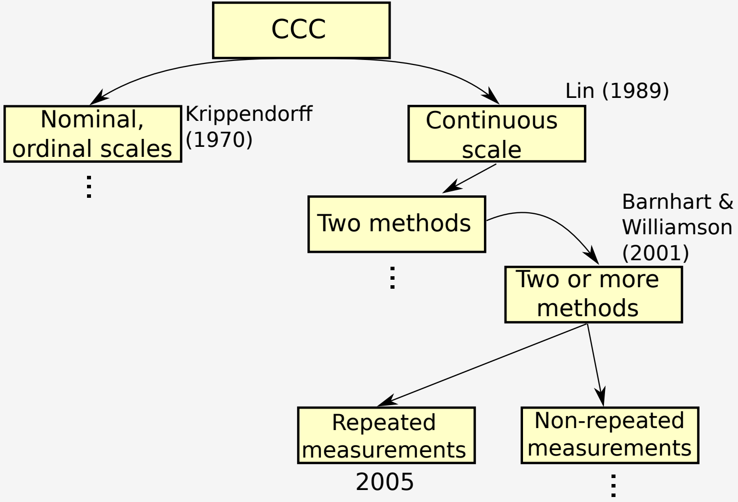
Concordance correlation coefficient

Bias correction factor (accuracy measure)

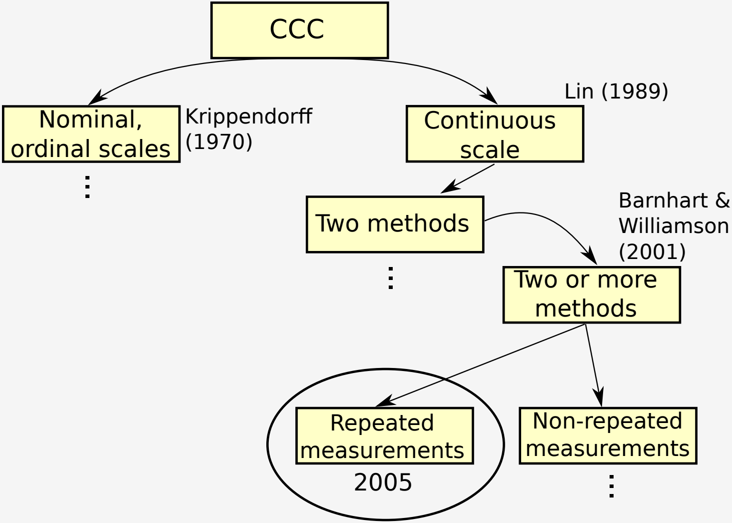
$C_b = \dfrac{2}{(v + 1/v + u^2)}$ is the accuracy measure (measures how far the best-fit line deviates from the 45° line through the origin); $v = \dfrac{\sigma_1}{\sigma_2}$ is the scale shift and $u = \dfrac{\mu_1 - \mu_2}{\sqrt{\sigma_1 \sigma_2}}$ is the location shift relative to the scale.



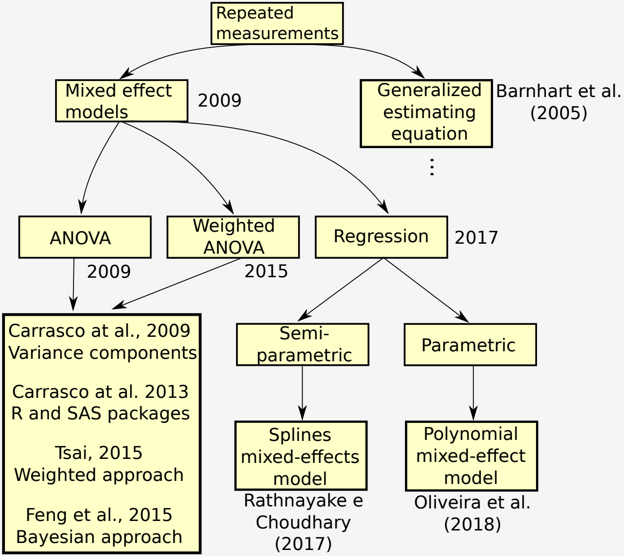
Literature review about CCC



Literature review about CCC



Literature review about CCC



Longitudinal Concordance Correlation Function

SpringerLink

Longitudinal Concordance Correlation Function Based on Variance Components: An Application in Fruit Color Analysis

Journal of Agricultural, Biological and Environmental Statistics

June 2018, Volume 23, Issue 2, pp 233–254 | Cite as

Article
First Online: 22 March 2018

Authors: Thiago de Paula Oliveira[1,2]
John Hinde[2]
Silvio Sandoval Zocchi[1]

1. University of São Paulo
2. National University of Ireland Galway

Article URL

Motivation: The papaya's hue colour analysis

- Global papaya producing:
  - India: 43,22% ($\approx$ 5,762,000)
  - Brazil: 11.13% ($\approx$ 1,484,000)
  - Mexico: 7,91% ($\approx$ 1,054,000)
- Papaya exports (global trade):
  - Mexico: 51.21%
  - Brazil: 14.02%
  - Guatemala: 9.97%
- Papayas are classified according to size and color;

Motivation: The papaya's hue colour analysis

- Global papaya producing:
  - India: 43,22% ($\approx$ 5,762,000)
  - Brazil: 11.13% ($\approx$ 1,484,000)
  - Mexico: 7,91% ($\approx$ 1,054,000)
- Papaya exports (global trade):
  - Mexico: 51.21%
  - Brazil: 14.02%
  - Guatemala: 9.97%
- Papayas are classified according to size and color;

Motivation: The papaya's hue colour analysis

- Global papaya producing:
  - India: 43,22% (≈ 5,762,000)
  - Brazil: 11.13% (≈ 1,484,000)
  - Mexico: 7,91% (≈ 1,054,000)
- Papaya exports (global trade):
  - Mexico: 51.21%
  - Brazil: 14.02%
  - Guatemala: 9.97%
- Papayas are classified according to size and color;

Methods for determining color

1. Colorimeter methodology

Konica Minolta (2002)

Zaragoza (2014)

- Very efficient when the fruit has uniform coloration, but it may be inefficient when the fruit does not have it.
- Usually, there are sampled some points in the equatorial region of the fruits;

Methods for determining color

1. Colorimeter methodology

Konica Minolta (2002)

Zaragoza (2014)

- Very efficient when the fruit has uniform coloration, but it may be inefficient when the fruit does not have it.
- Usually, there are sampled some points in the equatorial region of the fruits;

Methods for determining color

- The sample is observed from a very small area of the fruit's peel, then we may have the following problem:
  - The equatorial region may not be representative of the whole peel region;

Konica Minolta (2002)

Zaragoza (2014)

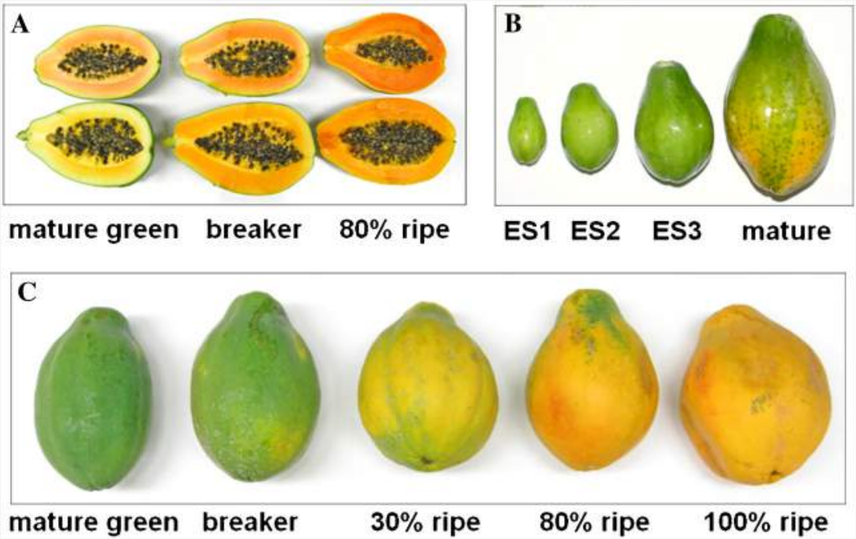

A
mature green
breaker
80% ripe
B
ES1
ES2
ES3
mature
C
mature green
breaker
30% ripe
80% ripe
100% ripe

Image analysis

2 Digital image analysis

- An alternative method was proposed to evaluate the color of several fruits based on image processing (Darrigues et al.,2008; Oliveira et al., 2017);
- Offers an objective measure for color and other physical factors;
- Most recent applications include classification and quality evaluation of several type of foods such as papaya, bananas, seeds, meats;

Image analysis

2 Digital image analysis

- An alternative method was proposed to evaluate the color of several fruits based on image processing (Darrigues et al.,2008; Oliveira et al., 2017);
- Offers an objective measure for color and other physical factors;
- Most recent applications include classification and quality evaluation of several type of foods such as papaya, bananas, seeds, meats;

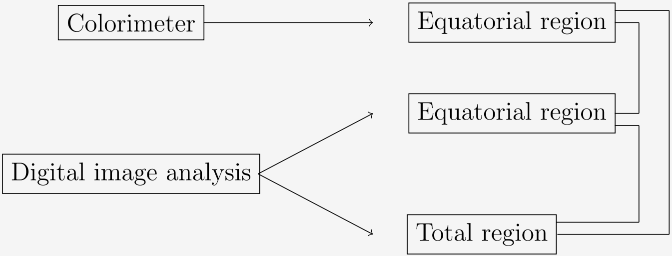
Image analysis

2 Digital image analysis

- An alternative method was proposed to evaluate the color of several fruits based on image processing (Darrigues et al.,2008; Oliveira et al., 2017);
- Offers an objective measure for color and other physical factors;
- Most recent applications include classification and quality evaluation of several type of foods such as papaya, bananas, seeds, meats;

Objective

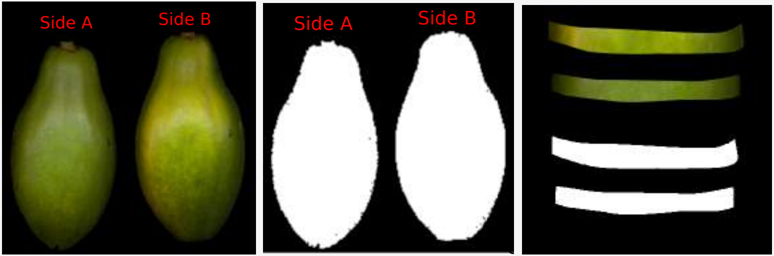
- Verify the agreement between digital image analysis and colorimeter methodologies over time, considering three main situations:



- Propose a longitudinal concordance correlation function (LCC) to verify the agreement among methods on same region, regions on same method, and methods on different regions;

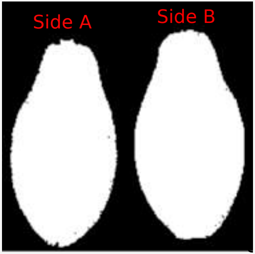
Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% ± 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



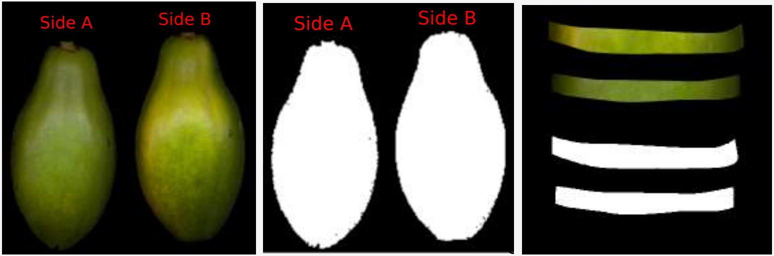
Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% ± 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



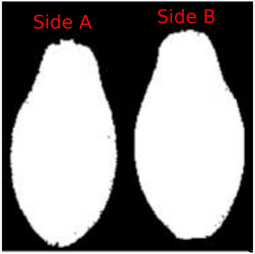
Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% ± 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



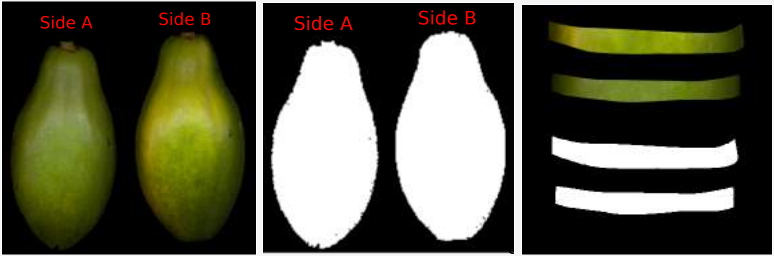
Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% $\pm$ 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



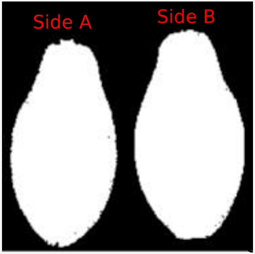
Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% $\pm$ 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



Experimental summary

- 20 papayas of cultivar Sunrise Solo;
- Harvested on the same day and classified in the same lot;
- Cooling chamber at 18°C and 80% ± 5% relative humidity;
- Observed daily
- Devices: Flat-bed scanner (HP Scanjet G2410) and a tristimulus colorimeter Minolta CR-300;
- The assessment period ranged from 11 to 14 days depending on the fruit.



Experimental summary

- Observed points on the peel:
  - Colorimeter: 4 points on equatorial region
  - Scanner:
    - 1000 points on equatorial region
    - 10,000 points on whole peel region

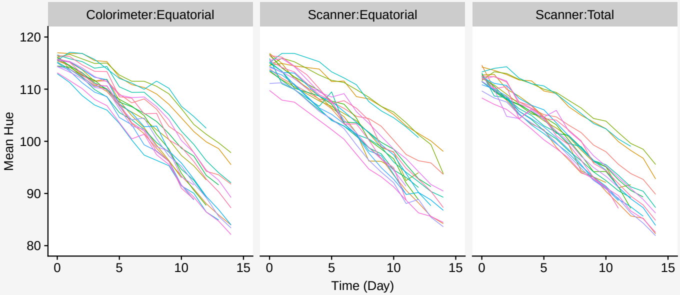


Figure: Individual profiles

Multiple mixed effects regression model

Definitions:

- y_{ijlk} is a realization of random variable Y_{ijlk}
- Indexes
  - $i = 1, 2, \ldots, N$ - individual;
  - $j = 1, 2, \ldots, m$ - method;
  - $l = 1, 2, \ldots, r$ - region;
  - $k = 1, 2, \ldots, n_i$ - time;

Multiple mixed effects regression model

- Assume that unobserved variables can differentially affect the response variable given the locality of the observed point on the peel.
- New variable A with mr categories given by the combination of region and method levels

$mr - 1$ associated dummy variables for fruit i at time k

$$\omega_{ick} = \begin{cases} 1, & \text{for category } c \text{ of variable } A \\ 0, & \text{otherwise} \end{cases}.$$

- Reference category mr corresponding to $\omega_{i1k} = \omega_{i2k} = \ldots = \omega_{i(mr-1)k} = 0$

Multiple mixed effects regression model

$$Y_{ijlk} = \sum_{h=0}^{p} \beta_{hjl} t_{ik}^{h} + \sum_{h=0}^{q} b_{hi} t_{ik}^{h} + \sum_{c=1}^{mr-1} \alpha_{ci} \omega_{ick} + \epsilon_{ijlk}$$

$$\boldsymbol{u}_i = \begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{\alpha}_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{G}_\alpha \end{bmatrix} \right) \quad \text{and} \tag{2}$$

$$\boldsymbol{\epsilon}_i \sim MVN\left(\mathbf{0}, \boldsymbol{R}_i\right),$$

- $q \leq p$
- Allows to model the correlation among repeated measures (constant or not constant over time)
- Remove the influence of the interaction among method, region, and fruit from the error term.

Heteroscedasticity assumption

- Subpopulation differences
  - equatorial and whole regions
  - different numbers of raw observations for calculating the mean hue
    - four points for colorimeter
    - 1000 for the scanner on the equatorial region
    - 10,000 for the scanner over the whole region

Variance model with different variances for each level of a stratification variable s, $s = 1, 2, \ldots, S$

$$\text{Var}\left(\epsilon_{ijlk}\right) = \sigma_\epsilon^2 \delta_{s_{ijlk}}^2.$$

General Variance function

$$\text{Var}\left(\epsilon_{ijlk}\right) = \sigma_\epsilon^2 g\left(t_{ik}, \boldsymbol{\delta}_{j'l'}\right)$$

General mixed effect model

General linear mixed model defined by Verbeke and Molenberghs (2000)

$$\begin{gathered} \boldsymbol{Y}_i = \boldsymbol{X}_i\left(t_{ik}\right)\boldsymbol{\beta} + \boldsymbol{Z}_i\left(t_{ik}\right)\boldsymbol{u}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{u}_i \sim MVN\left(\boldsymbol{0}, \boldsymbol{D}\right) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim MVN\left(\boldsymbol{0}, \boldsymbol{R}_i\right) \end{gathered} . \tag{3}$$

The longitudinal concordance correlation

- Under the model (2), we can define the LCC based on variance components for observations measured from different unique combinations of two factors at time t_{ik}.

LCC

$$\rho_{jl,\,j'l'}\left(t_{ik}\right) = \frac{2Cov\left(Y_{ijlk},\, Y_{ij'l'k}\right)}{Var\left(Y_{ijlk}\right) + Var\left(Y_{ij'l'k}\right) + \left[E\left(Y_{ijlk}\right) - E\left(Y_{ij'l'k}\right)\right]^2}. \quad (4)$$

The longitudinal concordance correlation

Let $\boldsymbol{z}_{ijlk}$ and $\boldsymbol{z}_{ij'l'k}$ be, respectively, rows of $\boldsymbol{Z}_i\left(t_{ik}\right)$:

$$\boldsymbol{z}_{ijlk} = \left(\boldsymbol{t}_{ik}, \boldsymbol{\omega}_{ik}\right) \quad \text{and} \quad \boldsymbol{z}_{ij'l'k} = \left(\boldsymbol{t}_{ik}, \boldsymbol{\omega}'_{ik}\right),$$

where $\boldsymbol{t}_{ik} = \left(t^0_{ik}, t^1_{ik}, \ldots, t^q_{ik}\right)$ and $\boldsymbol{\omega}_{ik} = \left(\omega_{i1k}, \omega_{i2k}, \ldots, \omega_{i(mr-1)k}\right)$

Covariance

$$Cov\left(Y_{ijlk}, Y_{ij'l'k}\right) = \boldsymbol{z}_{ijlk}\boldsymbol{D}\boldsymbol{z}^T_{ij'l'k} = \boldsymbol{t}_{ik}\boldsymbol{G}\boldsymbol{t}^T_{ik} + \boldsymbol{\omega}_{ik}\boldsymbol{G}_{\alpha}\boldsymbol{\omega}'^T_{ik}. \tag{5}$$

The longitudinal concordance correlation

Variance

$$\begin{cases} Var\left(Y_{ijlk}\right) = \boldsymbol{t}_{ik}\boldsymbol{G}\boldsymbol{t}_{ik}^{T} + \boldsymbol{\omega}_{ik}\boldsymbol{G}_{\alpha}\boldsymbol{\omega}_{ik}^{T} + \sigma_{\epsilon}^{2} g\left(t_{ik}, \boldsymbol{\delta}_{jl}\right) \\ Var\left(Y_{ij'l'k}\right) = \boldsymbol{t}_{ik}\boldsymbol{G}\boldsymbol{t}_{ik}^{T} + \boldsymbol{\omega}'_{ik}\boldsymbol{G}_{\alpha}\boldsymbol{\omega}'^{T}_{ik} + \sigma_{\epsilon}^{2} g\left(t_{ik}, \boldsymbol{\delta}_{j'l'}\right) \end{cases} . \tag{6}$$

Systematic differences

$$S_{jl,j'l'}\left(t_{ik}\right) = E\left(Y_{ijlk}\right) - E\left(Y_{ij'l'k}\right) = \mu_{jl}\left(t_{ik}\right) - \mu_{j'l'}\left(t_{ik}\right)$$

which reduces to

$$S_{jl,j'l'}\left(t_{ik}\right) = \boldsymbol{t}_{ik}\left(\boldsymbol{\beta}_{jl} - \boldsymbol{\beta}_{j'l'}\right), \text{ with } h = 1, 2, \ldots, p \text{ and } jl \neq j'l'. \tag{7}$$

The longitudinal concordance correlation

Variance

$$\begin{cases} Var\left(Y_{ijlk}\right) = \boldsymbol{t}_{ik}\boldsymbol{G}\boldsymbol{t}_{ik}^{T} + \boldsymbol{\omega}_{ik}\boldsymbol{G}_{\alpha}\boldsymbol{\omega}_{ik}^{T} + \sigma_{\epsilon}^{2} g\left(t_{ik}, \boldsymbol{\delta}_{jl}\right) \\ Var\left(Y_{ij'l'k}\right) = \boldsymbol{t}_{ik}\boldsymbol{G}\boldsymbol{t}_{ik}^{T} + \boldsymbol{\omega}'_{ik}\boldsymbol{G}_{\alpha}\boldsymbol{\omega}'^{T}_{ik} + \sigma_{\epsilon}^{2} g\left(t_{ik}, \boldsymbol{\delta}_{j'l'}\right) \end{cases} . \quad (6)$$

Systematic differences

$$S_{jl,\,j'l'}\left(t_{ik}\right) = E\left(Y_{ijlk}\right) - E\left(Y_{ij'l'k}\right) = \mu_{jl}\left(t_{ik}\right) - \mu_{j'l'}\left(t_{ik}\right)$$

which reduces to

$$S_{jl,\,j'l'}\left(t_{ik}\right) = \boldsymbol{t}_{ik}\left(\boldsymbol{\beta}_{jl} - \boldsymbol{\beta}_{j'l'}\right), \text{ with } h = 1, 2, \ldots, p \text{ and } jl \neq j'l'. \quad (7)$$

Definition of longitudinal concordance correlation function

Longitudinal concordance correlation function

$$\begin{aligned}\rho_{jl,\,j'l'}(t_k) &= \frac{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T}{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \frac{1}{2}\left\{\boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k^T + \boldsymbol{\omega}_k' \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T + \sigma_\epsilon^{2^*} + S^2_{jl,\,j'l'}(t_k)\right\}} \\ &= \rho^{(p)}_{jl,\,j'l'}(t_k)\, C_{jl,\,j'l'}(t_k) \end{aligned} \tag{8}$$

where, $\sigma_\epsilon^{2^*} = \sigma_\epsilon^2 \left[g\left(t_k, \boldsymbol{\delta}_{jl}\right) + g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)\right]$

- $\rho^{(p)}_{jl,\,j'l'}(t_k)$: the longitudinal Pearson correlation;
- $C_{jl,\,j'l'}(t_k)$: longitudinal accuracy;

Definition of longitudinal concordance correlation function

Longitudinal concordance correlation function

$$\rho_{jl,\,j'l'}(t_k) = \frac{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T}{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \frac{1}{2}\left\{\boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k^T + \boldsymbol{\omega}_k' \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T + \sigma_\epsilon^{2^*} + S^2_{jl,\,j'l'}(t_k)\right\}} = \rho^{(p)}_{jl,\,j'l'}(t_k)\, C_{jl,\,j'l'}(t_k) \quad (8)$$

where, $\sigma_\epsilon^{2^*} = \sigma_\epsilon^2 \left[g\left(t_k, \boldsymbol{\delta}_{jl}\right) + g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)\right]$

- $\rho^{(p)}_{jl,\,j'l'}(t_k)$: the longitudinal Pearson correlation;
- $C_{jl,\,j'l'}(t_k)$: longitudinal accuracy;

Longitudinal Pearson Correlation

Longitudinal Pearson Correlation

$$\rho^{(p)}_{jl,\, j'l'}\left(t_k\right) = \frac{\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T}{\sqrt{Var\left(Y_{ijlk}\right) Var\left(Y_{ij'l'k}\right)}}.$$

where

$$Var\left(Y_{ijlk}\right) = \left[\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k^T + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{jl}\right)\right]$$

and

$$Var\left(Y_{ij'l'k}\right) = \left[\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k' \boldsymbol{G}_\alpha \boldsymbol{\omega}_k'^T + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)\right]$$

- Measures how far each observation deviated from the best-fit line at a fixed time $t_k = t$

Longitudinal Accuracy

- Longitudinal bias correction factor (longitudinal accuracy) that measures how far the best-fit line deviates from the 45° line at a fixed time $t_k = t$

Longitudinal Accuracy

$$C_{jl,\, j'l'}\left(t_k\right) = \frac{2}{v_{jl,\, j'l'}\left(t_k\right) + \left[v_{jl,\, j'l'}\left(t_k\right)\right]^{-1} + u^2_{jl,\, j'l'}\left(t_k\right)}$$

where

The scale shift

$$v_{jl,\, j'l'}\left(t_k\right) = \sqrt{\frac{Var\left(Y_{ijlk}\right)}{Var\left(Y_{ij'l'k}\right)}} = \sqrt{\frac{t_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k^T + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{jl}\right)}{t_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}'_k \boldsymbol{G}_\alpha \boldsymbol{\omega}'^T_k + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)}}$$

Longitudinal Accuracy

- Longitudinal bias correction factor (longitudinal accuracy) that measures how far the best-fit line deviates from the 45° line at a fixed time $t_k = t$

Longitudinal Accuracy

$$C_{jl,\,j'l'}\left(t_k\right) = \frac{2}{v_{jl,\,j'l'}\left(t_k\right) + \left[v_{jl,\,j'l'}\left(t_k\right)\right]^{-1} + u^2_{jl,\,j'l'}\left(t_k\right)}$$

where

The scale shift

$$v_{jl,\,j'l'}\left(t_k\right) = \sqrt{\frac{Var\left(Y_{ijlk}\right)}{Var\left(Y_{ij'l'k}\right)}} = \sqrt{\frac{\boldsymbol{t}_k\boldsymbol{G}\boldsymbol{t}_k^T + \boldsymbol{\omega}_k\boldsymbol{G}_\alpha\boldsymbol{\omega}_k^T + \sigma^2_\epsilon g\left(t_k, \boldsymbol{\delta}_{jl}\right)}{\boldsymbol{t}_k\boldsymbol{G}\boldsymbol{t}_k^T + \boldsymbol{\omega}'_k\boldsymbol{G}_\alpha\boldsymbol{\omega}'^T_k + \sigma^2_\epsilon g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)}}$$

Longitudinal Accuracy

The location shift relative to the scale

$$u_{jl,\,j'l'}(t_k) = \frac{E\left(Y_{ijlk}\right) - E\left(Y_{ij'l'k}\right)}{\left[Var\left(Y_{ijlk}\right) Var\left(Y_{ij'l'k}\right)\right]^{\frac{1}{4}}} = \frac{\boldsymbol{t}_k\left(\boldsymbol{\beta}_{jl} - \boldsymbol{\beta}_{j'l'}\right)}{\left[Var\left(Y_{ijlk}\right) Var\left(Y_{ij'l'k}\right)\right]^{\frac{1}{4}}}$$

where

$$Var\left(Y_{ijlk}\right) = \left[\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}_k \boldsymbol{G}_\alpha \boldsymbol{\omega}_k^T + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{jl}\right)\right]$$

and

$$Var\left(Y_{ij'l'k}\right) = \left[\boldsymbol{t}_k \boldsymbol{G} \boldsymbol{t}_k^T + \boldsymbol{\omega}'_k \boldsymbol{G}_\alpha \boldsymbol{\omega}'^T_k + \sigma_\epsilon^2 g\left(t_k, \boldsymbol{\delta}_{j'l'}\right)\right]$$

Estimation of the LCC using variance components

- We use the restricted maximum likelihood approach

The log-likelihood function to maximize is proportional to

$$l_R(\boldsymbol{\beta}, \boldsymbol{\psi}_u, \boldsymbol{\psi}_\epsilon; \boldsymbol{y}) \propto -\frac{1}{2}\left\{\log\left|\boldsymbol{V}\right| + \left|\boldsymbol{X}(t_{ik})^T \boldsymbol{V}^{-1}\boldsymbol{X}(t_{ik})\right| + \boldsymbol{r}^T\boldsymbol{V}^{-1}\boldsymbol{r}\right\},$$

- $\boldsymbol{\psi}_u$: vector of variance components of the $\boldsymbol{D}$ matrix;
- $\boldsymbol{\psi}_\epsilon$: vector of variance components of the $\boldsymbol{R}_i$ matrix;
- $r = (\boldsymbol{y} - \boldsymbol{X}(t_{ik})\boldsymbol{\beta})$: marginal residual

- $\rho_{jl,j'l'}(t_{ik})$ can be estimated replacing $\boldsymbol{\beta}$, $\boldsymbol{\psi}_u$, and $\boldsymbol{\psi}_\epsilon$ by their REML estimates.

Non-parametric confidence intervals

- We consider a simple case-resampling bootstrap;
- Generate B pseudo-samples by resampling from the data;
- Refit the model in order to obtain B sets of estimates for all parameters of the LCC;
- Fisher Z-transformation $\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)\right)$

Results

- Equatorial region by the scanner and colorimeter

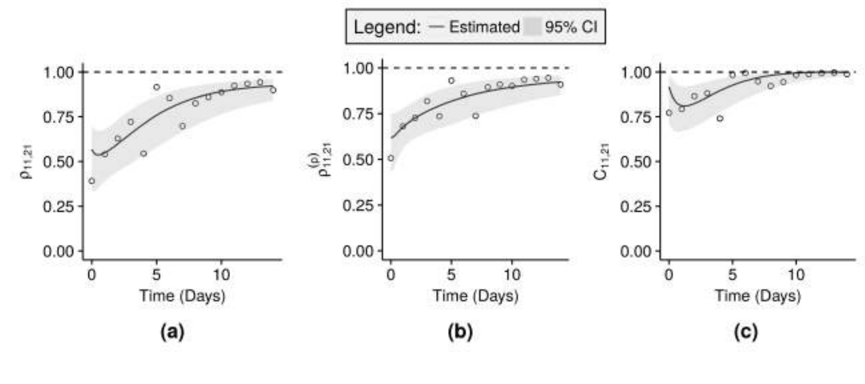


Figure: Estimate and 95% confidence interval (CI) for the longitudinal concordance correlation (a); longitudinal Pearson correlation (b); and longitudinal accuracy (c)

Results

- Equatorial region by the colorimeter and whole region by the scanner

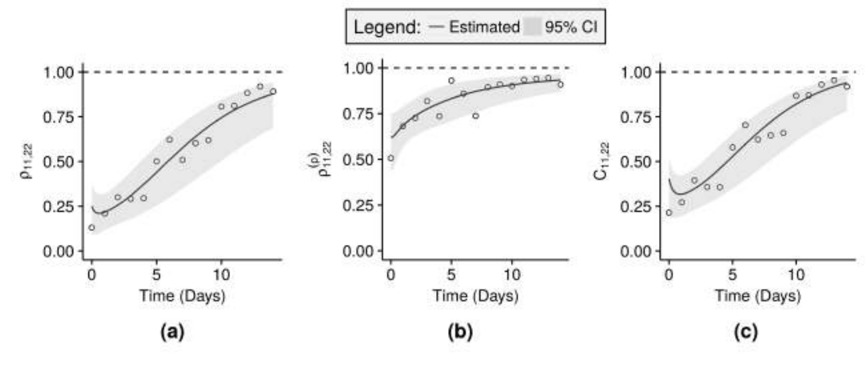


Figure: Estimate and 95% confidence interval (CI) for the longitudinal concordance correlation (a); longitudinal Pearson correlation (b); and longitudinal accuracy (c)

Results

- Equatorial and whole regions by the scanner

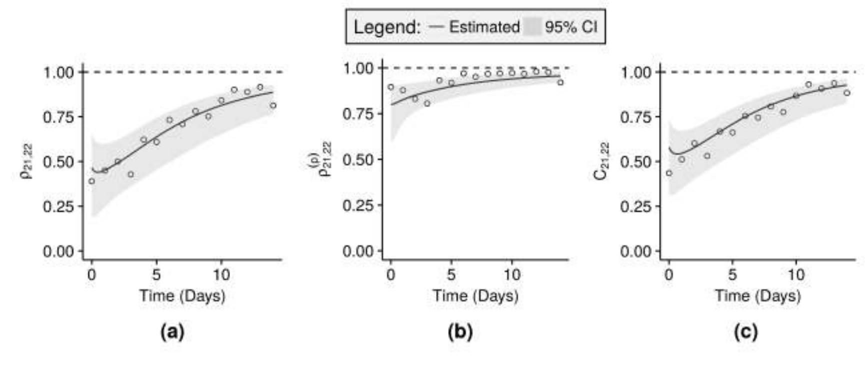


Figure: Estimate and 95% confidence interval (CI) for the longitudinal concordance correlation (a); longitudinal Pearson correlation (b); and longitudinal accuracy (c)

Conclusions

- LCC, as well as LPC and LA, showed that sample points only on the equatorial region were not representative of the whole peel region
- Image analysis of the whole peel region should be used to compute the mean hue
- LA between observations measured by the colorimeter and scanner on the equatorial region suggested that the topography and curved surface of papaya fruit did not affect the mean hue obtained by the scanner

Conclusions

- LCC, as well as LPC and LA, showed that sample points only on the equatorial region were not representative of the whole peel region
- Image analysis of the whole peel region should be used to compute the mean hue
- LA between observations measured by the colorimeter and scanner on the equatorial region suggested that the topography and curved surface of papaya fruit did not affect the mean hue obtained by the scanner

Conclusions

- LCC, as well as LPC and LA, showed that sample points only on the equatorial region were not representative of the whole peel region
- Image analysis of the whole peel region should be used to compute the mean hue
- LA between observations measured by the colorimeter and scanner on the equatorial region suggested that the topography and curved surface of papaya fruit did not affect the mean hue obtained by the scanner

The `lcc` package

- The version 1.0.1 is available to download in the CRAN (link lcc)
- The `lcc()` function:
  - fitted values for the LCC, LPC, and LA statistics
  - Non-parametric bootstrap confidence intervals
- `summary()` function:
  - method implemented for 'lcc' objects
  - `summary(obj, type = "model")` returns the mixed effect regression model
  - `summary(obj, type = "lcc")` returns fitted values and confidence intervals
- `lccPlot()` function:
  - returns a ggplot based graphics
  - fitted values are joined by lines while sampled observations are represented by circles

The `lcc` package

- The version 1.0.1 is available to download in the CRAN (link lcc)
- The `lcc()` function:
  - fitted values for the LCC, LPC, and LA statistics
  - Non-parametric bootstrap confidence intervals
- `summary()` function:
  - method implemented for 'lcc' objects
  - `summary(obj, type = "model")` returns the mixed effect regression model
  - `summary(obj, type = "lcc")` returns fitted values and confidence intervals
- `lccPlot()` function:
  - returns a ggplot based graphics
  - fitted values are joined by lines while sampled observations are represented by circles

The `lcc` package

- The version 1.0.1 is available to download in the CRAN (link lcc)
- The `lcc()` function:
  - fitted values for the LCC, LPC, and LA statistics
  - Non-parametric bootstrap confidence intervals
- `summary()` function:
  - method implemented for 'lcc' objects
  - `summary(obj, type = "model")` returns the mixed effect regression model
  - `summary(obj, type = "lcc")` returns fitted values and confidence intervals
- `lccPlot()` function:
  - returns a ggplot based graphics
  - fitted values are joined by lines while sampled observations are represented by circles

Example: Percentage body fat dataset

- Observational study conducted by the Penn State Young Women's Health Study (Loyd et. al, 1998)
- Response variable: Percentage body fat
- Methods: Skinfold calipers and Dual-energy x-ray absorptiometry (DEXA)
- Individual: cohort of 82 adolescent white females attending public school in Pennsylvania

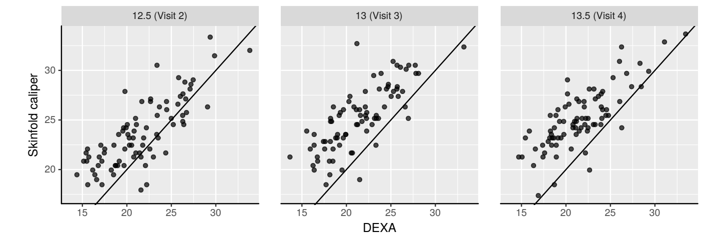


Figure: Scatter plot of body fat data, where the panels represent visits

Model

$$y_{ijk} = \beta_{0j} + b_{0i} + (\beta_{1j} + b_{1i})\, t_k + \epsilon_{ijk}$$
$$\boldsymbol{b} = [b_{0i}, b_{1i}]^T \sim N_2(\mathbf{0}, \boldsymbol{G}) \quad \text{and} \quad \epsilon_{ijk} \sim N\left(0, \sigma^2_\epsilon\right), \tag{9}$$

lcc function

	SUBJECT	TIME	BF	MET
1	101	6	21.68	1
2	102	6	18.47	1
3	103	6	21.68	1
4	105	6	23.19	1

```
R> fm1 <- lcc(dataset = bfat, subject = "SUBJECT", resp = "BF",
              method = "MET", time = "TIME", qf = 1, qr = 1,
              components = TRUE, ci = TRUE, nboot = 5000)
```

summary function

```
R > summary(fm1)
$Summary.lcc$fitted$LCC
   Time  LCC: DEXA vs. Skinfold     Lower      Upper
1     6            0.6653516    0.5703188  0.7376588
2    12            0.5589258    0.4512423  0.6434142
3    18            0.4588008    0.3332018  0.5601570
$Summary.lcc$fitted$LPC
   Time   LPC: DEXA vs. Skinfold     Lower      Upper
1     6            0.8065578    0.7423290  0.8551368
2    12            0.7826493    0.7098433  0.8373112
3    18            0.7620551    0.6672221  0.8300010
$Summary.lcc$fitted$LA
   Time   LA: DEXA vs. Skinfold     Lower      Upper
1     6           0.8249273     0.7373466  0.8860078
2    12           0.7141458     0.6155902  0.7897535
3    18           0.6020573     0.4893392  0.6950811
```

plot function

```
R> lccPlot(fm1, type="lcc")
R> lccPlot(fm1, type="lpc")
R> lccPlot(fm1, type="la")
```

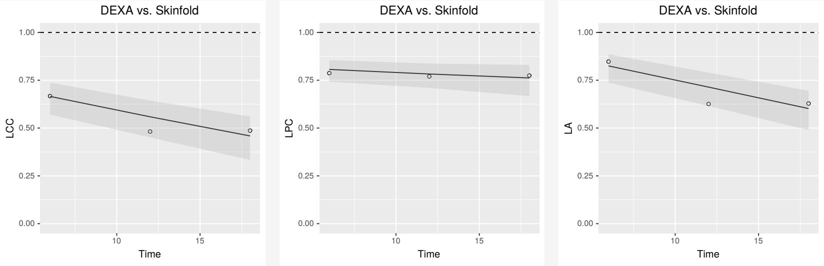


Figure: Estimate and 95% bootstrap confidence interval for the (a) LCC; (b) LPC; and (c) LA between percentage body fat measured on adolescent girls by skinfold caliper and DEXA. In addition, points represent the sample CCC, sample Pearson correlation, and sample accuracy, respectively

References

Carrasco, J. L., King, T. S., and Chinchilli, V. M. The concordance correlation coefficient for repeated measures estimated by variance components. **Journal of Biopharmaceutical Statistics**, 19(1):90–105, 2009.

Carrasco, J. L., Phillips, B. R., Puig-Martinez, J., King, T. S., and Chinchilli, V. M. Estimation of the concor- dance correlation coefficient for repeated measures using SAS and R. **Computer Methods and Programs in Biomedicine**, 109:293–304, 2013.

Feng, D.; Svetnik, V.; Coimbra, A.; Baumgartner, R. A comparison of confidence interval methods for the concordance correlation coefficient and intraclass correlation coefficient with small number of raters. **Journal of biopharmaceutical statistics**, 24(2):272-293, 2015.

Krippendorff, K. Bivariate agreement coefficients for reliability of data. **Sociological Methodology**, 2:139–150, 1970

Lin, L. I. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, 45(1):255–268, 1989.

References

Oliveira, T.P.; Hinde, J.; Zocchi, S.S. Longitudinal concordance correlation function based on variance components: an application in fruit color analysis. **Journal of Agricultural, Biological, and Environmental Statistics**, 23(2): 233-254, 2018.

R core Team. The R environment, 2015. URL https://www.r-project.org/.

Rathnayake, L. N. and Choudhary, P. K. Semiparametric modeling and analysis of longitudinal method comparison data. **Statistics in Medicine**, 36(13):2003–2015, 2017.

Tsai, M. Comparison of concordance correlation coefficient via variance components, generalized estimating equations and weighted approaches with model selection. **Computational Statistics and Data Analysis**, 82(1):47-58, 2015.

Verbeke, G. and Molenberghs, G. **Linear mixed models for longitudinal data**. Springer, New York, 2000.