Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Notes and Comments

II Production Unknown: a R package for application of the percentage of importance indice-production unknown

G. L. Demolin-Leite[a]* and A. M. Azevedo[a]

[a]Universidade Federal de Minas Gerais – UFMG, Instituto de Ciências Agrárias – ICA, Montes Claros, MG, Brasil

The Importance Indice (*I.I.*) can determine the loss (*L.S.*) and solution sources (*S.S.*) for a system in certain knowledge areas (e.g., agronomy), when production (e.g., fruits) is known (Demolin-Leite, 2021, 2024). However, the final production of the system is not always known or is difficult to determine (e.g., degraded area recovery). A derivation of the *I.I.* is the percentage of Importance Indice-Production Unknown (*% I.I.-PU*) that can detect the loss or solution sources when production is unknown for the system (Demolin-Leite, 2022). This new index can help in monitoring degraded area recovery. This index and its derivations (e.g., reduction of the total *n.* of *L.S.* (*R.L.S*)/total n. of the *S.S*) were obtained using the statistical programs Biodiversity Professional program, version 2 (Krebs, 1989) – for chi-square test - and System for Analysis Statistics and Genetics, version 9.1 (UFV, 2007) – for simple regression analysis (e.g., damage)-, and also part of the calculations using an Excel datasheet (e.g., percentage of *R.L.S.* per *S.S.*). However, the transfer of information from the data obtained via the statistical programs mentioned above, as well as the calculations performed using the Excel datasheet, in addition to being labor intensive, could incur mathematical errors due to the volume of equations and data. For this purpose, a package and its manual were developed, via the R program, to perform the statistics and calculations necessary to obtain the *%I.I.P.U.* and its derivations (Demolin-Leite and Azevedo, 2022). This study aimed to demonstrate to use of the R-Package 'IIProductionUnknown' (Demolin-Leite and Azevedo, 2022) using adapted published data (simplified) (see Demolin-Leite, 2022) about those obtained with the statistical programs mentioned above. The package is available on Cran's platform (https://cran.r-project.org/web/packages/IIProductionUnknown/IIProductionUnknown.pdf).

Percentage of Importance Index-Production Unknown (*% I.I.-PU*) (Demolin-Leite, 2022) is: $\% I.I.\text{-}PU = [(ks_1 \times c_1 \times ds_1)/\Sigma(ks_1 \times c_1 \times ds_1)+(ks_2 \times c_2 \times ds_2)+(ks_n \times c_n \times ds_n)] \times 100$ (Demolin-Leite, 2022), where,

i) the *key source* (*ks*) is: *ks* = damage (non-percentage) (*Da.*)/total *n* of the *L.S.* on the samples or *ks* = reduction of the total *n.* of *L.S.* (*R.L.S*)/total n. of the *S.S* on the samples (Demolin-Leite, 2022). Where *Da.* or *R.L.S.* = $R^2 \times (1 - P)$, when it is of the first degree, or $((R^2 \times (1 - P)) \times (\beta_2/\beta_1)$, when it is of the second degree, where $\mathbf{R^2}$ = determination coefficient and ***P***= significance of ANOVA, β_1 = regression coefficient, and β_2 = regression coefficient (variable²), of the simple regression equation of the loss source (*L.S.*) or solution source (*S.S.*) (Table 1) (Demolin-Leite, 2022). When it is not possible to separate the *Da.* between two or more *L.S.*, divide the *Da.* among the *L.S.* as a proportion of their respective "total *n*". *Da.* = 0 when *Da.* was non-significant for damage or non-detected by *L.S.* on the system (Demolin-Leite, 2022). When an *S.S.* operates in more than one *L.S.*, that caused damage, its *ks* are summed. *R.L.S.* = 0 when *Da.* by *L.S.* or *R.L.S.* was non-significant for damage by *L.S.* or reduced *L.S.* by *S.S.* on the system (Demolin-Leite, 2022).

ii) *c* (*constancy*) = Σ of occurrence of *L.S.* or *S.S.* on samples, where absence = 0 or presence = 1 (Demolin-Leite, 2021). And

iii) *ds* (*distribution source*) = 1 - ***P*** of the chi-square test of *L.S.* or *S.S.* on the samples (Table 1) (Demolin-Leite, 2021). Counts (non-frequency) of *L.S.* or *S.S.* are used to perform the chi-square test.

The data above (Chart 1) are obtained by R-package (Demolin-Leite and Azevedo, 2022):

The percentage of *R.L.S.* per *S.S.* (*%R.L.S.S.S.*) = (*R.L.S.S.S.*/total *n* of the *L.S.* – abundance or damage) x 100, where *R.L.S.S.S.* = *R.L.S.* x total *n* of the *S.S.*, with the *R.L.S.* not being summed in this case (Demolin-Leite, 2022).

The data above (Chart 2) are obtained by R-package (Demolin-Leite and Azevedo, 2022):

The loss sources (*L.S.*, e.g., insect pests) $L.S._4$ and $L.S._2$ showed the highest *% I.I.* (41.73 and 39.98%, respectively) on 48 samples. The effective solution sources (*S.S.*, e.g., natural enemies) $S.S._4$ and $S.S._1$ showed the highest *% I.I.* (52.46 and 47.54%, respectively) on 48 samples (Table 2). The number of $L.S._3$ was reduced per number of $S.S._4$ (13.85%) and that of $L.S._4$. that of $S.S._1$ (1.02%). However, the number of $L.S._2$ was increased per number of $S.S._3$ (2.07%) and that of $L.S._4$. that of $S.S._2$ (93.14%) on the samples. The final balance was negative on the system, with an increase of lost sources of 80.34% in these saplings. The $L.S._4$. damage was reduced per number of $S.S._1$ (6.03%); but the $L.S._2$ damage was increased per number of $S.S._3$ (29.67%), totaling 23.64% of the increase by lost source damages on the samples (Table 3).

*e-mail: germano.demolin@gmail.com
Received: November 19, 2022 – Accepted: March 15, 2023

Table 1. Aggregated, regular, or random distribution of the loss (*L.S.*) or solution sources (*S.S.*); and simple regression equations with their coefficients of determination (R^2), significance (*P*), and *F* of the analysis of variance (ANOVA) of damage per *L.S.* and reduction or increase of *L.S.* (abundance or damage) per *S.S.* on 48 samples.

L.S.	**Chi-square test – R program function: ChisqTest_Distribution**			
	Variance	**Mean**	***P***	**Distribution**
$L.S._1$	0.60	0.69	0.71	Random
$L.S._2$	66.97	2.58	0.00	Aggregated
$L.S._3$	0.20	0.19	0.37	Random
$L.S._4$	8.14	0.77	0.00	Aggregated
$L.S._5$	1.98	1.02	0.00	Aggregated
S.S.				
$S.S._1$	0.14	0.10	0.01	Random
$S.S._2$	113.50	5.17	0.00	Aggregated
$S.S._3$	0.84	0.19	0.00	Aggregated
$S.S._4$	0.28	0.25	0.03	Random
$S.S._5$	1.16	0.90	0.01	Random

Simple regression analysis – R program function: Loss Source and Effectiveness Of Solution		**ANOVA**	
	R^2	***F***	***P***
Defoliation=3.83+0.21x$L.S._5$	0.09	4.43	0.0409
Damage=-0.03+0.08x$L.S._2$	0.98	1881.60	0.0000
Damage=-0.0005+0.19x$L.S._4$	0.99	3217.00	0.0000
$L.S._2$ damage=0.11+0.42x$S.S._3$	0.30	19.41	0.0001
$L.S._4$ damage=0.09+1.86x$S.S._1$-0.96x$S.S._1^2$	0.17	4.53	0.0161
$L.S._2$=1.69+4.78x$S.S._3$	0.29	18.32	0.0001
$L.S._3$=0.11+1.09x$S.S._4$-0.57x$S.S._4^2$	0.20	5.61	0.0067
$L.S._4$=0.50+9.25x$S.S._1$-4.75x$S.S._1^2$	0.15	3.98	0.0256
$L.S._4$=0.25+0.10x$S.S._2$	0.14	7.50	0.0088

Chart 1. Steps used to obtain the above data.

library(IIProductionUnknown)
data("DataLossSource")
Description: an example with data from loss sources with five loss sources, one in each column.
ChisqTest_Distribution(DataLossSource)
Description: indicates the distribution of loss sources: aggregate, random, or regular. It is a matrix object containing data from loss sources.
data("DataSolutionSource")
Description: an example with data from solution sources with five solution sources, one in each column.
ChisqTest_Distribution(DataSolutionSource)
Description: indicates the distribution of solution sources: aggregate, random, or regular. It is a matrix object containing data from solution sources.

These numbers obtained using the R-Package 'IIProductionUnknown' (Demolin-Leite and Azevedo, 2022), were the same as those obtained with the adapted data (reduced) of the published paper (Demolin-Leite, 2022). However, the R-Package 'IIProductionUnknown' (Demolin-Leite and Azevedo, 2022) was faster, more practical, and safer way than those obtained previously via the statistical programs and Excel datasheet mentioned above. This new index can help, as an example, in monitoring degraded area recovery.

Chart 2. Steps used to obtain the above data.

data("DataDefoliation")

```
data("DataDamage")
DataResult<-cbind(DataDefoliation,DataDamage$D.L.S.2,DataDefoliation,
DataDamage$D.L.S.4,DataDefoliation)
ResultLossSource<-LossSource(DataLoss = DataLossSource,DataResult =DataResult,Cols=c(1,3,5),verbose=TRUE)
EOS<-EffectivenessOfSolution(DataLossSource =DataLossSource,
DataSolutionSource =DataSolutionSource,ResultLossSource = ResultLossSource)
EOS
ID<-SelectEffectivenessOfSolution(EOS)
ID
ResultSolutionSource<-SolutionSource(SolutionData =DataSolutionSource,Production =DataResult,
EffectivenessOfSolution=EOS, Id = ID,Verbose = TRUE)
ResultSolutionSource
ReductionAbundance(ResultSolutionSource,ResultLossSource,
EffectivenessOfSolution=EOS)
EOSDamage<-EffectivenessOfSolution(DataLossSource =DataDamage,
DataSolutionSource =DataSolutionSource,ResultLossSource = NULL)
EOSDamage
ReductionDamage(ResultSolutionSource,LossSource=DataDamage,
EffectivenessOfSolution=EOSDamage)
```

Table 2. Total number (*n*), damage (*Da.*) or reduction of *L.S.* (*R.L.S.*), key-source (*ks*), constancy (*c*), distribution source (*ds*), number of importance indices (*n. I.I.*), the sum of *n. I.I.-PU* (Σ *n. I.I.*), and percentage of *I.I.* by loss source (*L.S.*) on 48 samples.

Loss sources – R program function: LossSource								
L.S.	***n***	***Da.***	***ks***	***c***	***ds***	***n.I.I.***	**Σ*n.I.I.***	***%I.I.***
L.S.$_1$	33	0.0147	0.0005	26	0.29	0.0033	0.2560	1.30
L.S.$_2$	124	0.9761	0.0079	13	1.00	0.1023	0.2560	39.98
L.S.$_3$	9	0.0040	0.0004	8	0.63	0.0023	0.2560	0.88
L.S.$_4$	37	0.9882	0.0267	4	1.00	0.1068	0.2560	41.73
L.S.$_5$	49	0.0842	0.0017	24	1.00	0.0412	0.2560	16.11
Solution sources – R program function: SolutionSource								
S.S.	***n***	***R.L.S.***	***ks***	***c***	***ds***	***n.I.I.***	**Σ*n.I.I.***	***%I.I.***
S.S.$_1$	5	0.07520	0.01500	4	0.93	0.0561	0.1179	47.54
S.S.$_2$	248	0.00000	0.00000	27	1.00	0.0000	0.1179	0.00
S.S.$_3$	9	0.00000	0.00000	3	1.00	0.0000	0.1179	0.00
S.S.$_4$	12	0.10389	0.00866	10	0.72	0.0619	0.1179	52.46
S.S.$_5$	43	0	0	27	0.92	0.0000	0.1179	0.00

Table 3. Percentage of reduction in abundance and or damage (*%R.*) of loss source (*L.S.*) per solution source (*S.S.*), sum (Σ), and a total of Σ of *R.L.S.* (T.Σ) on 48 samples.

	%R.L.S.S.S.- abundance **– R program function: ReductionAbundance**		
	L.S.		
S.S.	***L.S.*$_2$**	***L.S.*$_3$**	***L.S.*$_4$**
S.S.$_1$	---	---	1.02
S.S.$_2$	---	---	-93.14
S.S.$_3$	-2.07	---	---
S.S.$_4$	---	13.85	
Σ	**-2.07**	**13.85**	**-92.12**
***T.Σ**	**-80.34**	---	---
	%R.L.S.S.S.- damage **– R program function: ReductionDamage**		
	L.S.		
S.S.	***L.S.*$_2$**	***L.S.*$_3$**	***L.S.*$_4$**
S.S.$_1$	---	---	6.03
S.S.$_3$	-29.67	---	---
Σ	**-29.67**	---	**6.03**
***T.Σ**	**-23.64**	---	---

*T.Σ = total sum.

References

DEMOLIN-LEITE, G.L., 2021 [viewed 2 May 2018]. Importance indice: loss estimates and solution effectiveness on production. *Canadian Journal of Agricultural Science* [online], vol. 55, no. 2, pp. 1-7. Available from: http://scielo.sld.cu/pdf/cjas/v55n2/2079-3480-cjas-55-02-e10.pdf

DEMOLIN-LEITE, G.L., 2022. Percentage of importance indice-production unknown: loss and solution sources identification on system. *Brazilian Journal of Biology = Revista Brasileira de Biologia*, vol. 84, pp. e253218. http://dx.doi.org/10.1590/1519-6984.253218. PMid:35019097.

DEMOLIN-LEITE, G.L., 2024. Do arthropods and diseases affect the production of fruits on *Caryocar brasiliense* Camb. (Malpighiales: caryocaraceae)? *Brazilian Journal of Biology = Revista Brasileira de Biologia*, vol. 84, pp. e253215. http://dx.doi.org/10.1590/1519-6984.253215.

DEMOLIN-LEITE, G.L. and AZEVEDO, A.M., 2022 [viewed 2 May 2018]. *Package 'IIProductionUnknown': analyzing data through of percentage of importance indice (production unknown) and its derivations: manual package* [online]. p. 1-18. Available from: https://cran.r-project.org/web/packages/IIProductionUnknown/IIProductionUnknown.pdf

KREBS, C.J., 1989 [viewed 2 May 2018]. *Bray-Curtis cluster analysis* [online]. Available from: http://biodiversity-pro.software.informer.com/

UNIVERSIDADE FEDERAL DE VIÇOSA – UFV, 2007. *Sistema para Análises Estatísticas e Genéticas (SAEG) version 9.1* [online]. Available from: http://arquivo.ufv.br/saeg/(accessed 30 june 2018).