

BPMSG Diversity Calculator

Author: Klaus D. Goepel – <http://bpmsg.com>

Overview

The diversity excel template allows to calculate α -, β - and γ -diversity for a set samples (input data), and analyze similarities between the samples based on β -diversity.

The template works under Windows OS and Excel 2010 (xlsx extension). No macros or links to external workbooks are necessary. The workbook consists of an input worksheet for data samples, a calculation worksheet where all necessary calculations are done a worksheet "beta" displaying the results.

Limitations

- Maximal number of data samples: 24
- Maximal number of categories/classes per data sample: 20

Results

Following data will be displayed:

For all samples:

- a) **Shannon Entropy H** (natural logarithm) α -, β - and γ , and corresponding Hill numbers 1D

Diversity (Shannon - ln)		1D
α -Diversity	1,557	4,745
γ -Diversity	1,775	5,903
β -Diversity	0,218	1,244

- b) **Homogeneity Measures**

Homogeneity Measures	
MacArthur M	0,8038
rel. Homogeneity 1S	78,9%
AHP consensus S^*	61,1%

Mac Arthur $M = 1/^1D_\beta$ homogeneity indicator

relative Homogeneity $^1S = (M - 1/N) / (1 - 1/N)$; N number of samples (between 0 and 100%)

AHP group consensus S^* : relative homogeneity corrected by AHP scale limitations

- c) **AHP max γ and min β true diversity based on the max score of the AHP scale**

AHP	
scale max	9
D_γ -max	6,88
D_α -min	3,41

Table 1: Shannon α -Entropy, Equitability, Simpson Dominance, Gini-Simpson index and Hill numbers

Sample	No	Shannon α -Entropy	Equi- tability	Simpson Dominance	Gini- Simpson	Hill Numbers 1D	2D
sample-1	1	1,6216	91%	0,2270	77,3%	5,06	4,41
sample-2	2	1,4544	81%	0,2873	71,3%	4,28	3,48
sample-3	3	1,6684	93%	0,2022	79,8%	5,30	4,95
sample-4	4	1,5728	88%	0,2309	76,9%	4,82	4,33

Table 2: will show the top 20 pairs of most similar samples

No	S	Sample (1)	Sample (2)	α	β
	100%				
1	97%	sample-4	sample-10	1,63	1,63
2	97%	sample-7	sample-14	1,47	1,47
3	96%	sample-9	sample-13	1,45	1,45
4	96%	sample-3	sample-14	1,64	1,64

On page 2 the matrix of pairs of data samples is displayed. It shows either α -, β - or γ -diversity (Shannon) of each pair of data samples, Horn Index or AHP consensus indicator.

Shannon α -Diversity																										
		2005	2006	Q3 2007	Q1 2008	Q4 2007	Q2 2008	Q2 2009	Q3 2008	Q1 2009	Q4 2008	Q3 2009	Q4 2009	Q1 2010	Q2 2010	Q3 2010	Q4 2010	Q1 2011	Q2 2011	Q3 2011	Q4 2011	Q4 2012	Q1 2012	Q3 2012	Q2 2012	
		1	2	3	5	4	6	10	7	9	8	11	12	13	14	15	16	17	18	19	20	24	21	23	22	
		1,55	1,34	1,36	1,49	1,56	1,55	1,53	1,56	1,57	1,57	1,6	1,59	1,59	1,58	1,57	1,59	1,58	1,57	1,56	1,6	1,6	1,59	1,6	1,57	1,55
4,72		3,8	3,9	4,4	4,8	4,7	4,6	4,8	4,8	4,8	4,9	4,9	4,9	4,8	4,8	4,9	4,8	4,8	4,7	5,0	5,0	4,9	4,9	4,8	4,7	
2005 2006 Q3 2007 Q1 2008 Q4 2007 Q2 2008 Q2 2009 Q3 2008 Q1 2009 Q4 2008 Q3 2009 Q4 2009 Q1 2010 Q2 2010 Q3 2010 Q4 2010 Q1 2011 Q2 2011 Q3 2011 Q4 2011 Q4 2012 Q1 2012 Q2 2012	1	0,00	0,00	0,04	0,05	0,06	0,06	0,08	0,08	0,07	0,07	0,09	0,09	0,11	0,12	0,11	0,12	0,13	0,14	0,10	0,10	0,11	0,10	0,12	0,14	
	2	0,00	0,00	0,03	0,05	0,05	0,06	0,08	0,07	0,07	0,07	0,09	0,08	0,10	0,11	0,11	0,12	0,12	0,13	0,10	0,10	0,10	0,10	0,12	0,14	
	3	0,04	0,03	0,00	0,00	0,01	0,01	0,02	0,03	0,03	0,03	0,03	0,03	0,04	0,04	0,05	0,05	0,06	0,06	0,05	0,05	0,06	0,05	0,07	0,08	
	5	0,05	0,05	0,01	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,02	0,02	0,02	0,02	0,03	0,02	0,02	0,03	0,03	0,04	0,04
	4	0,06	0,05	0,01	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,03	0,02	0,02	0,03	0,03	0,05	0,04
	6	0,06	0,06	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02	0,02	0,03	0,02	0,03	0,04	0,04	0,05	0,04
	10	0,08	0,08	0,03	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,02	0,02	0,03	0,04	0,02
	7	0,08	0,07	0,03	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,02	0,02	0,03	0,02
	9	0,07	0,07	0,03	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,02	0,02	0,03	0,03	0,02
	8	0,07	0,07	0,03	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,02	0,02
	11	0,09	0,09	0,03	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,01	0,01	0,01	0,02	0,01
	12	0,09	0,08	0,03	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,02	0,02
	13	0,11	0,10	0,04	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,03	0,01	
	14	0,12	0,11	0,04	0,02	0,01	0,02	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,03	0,01	
	15	0,11	0,11	0,05	0,02	0,02	0,02	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,01	
	16	0,12	0,12	0,05	0,02	0,02	0,02	0,01	0,01	0,01	0,01	0,00	0,01	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,01
	17	0,13	0,12	0,06	0,02	0,02	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,01
	18	0,14	0,13	0,06	0,03	0,03	0,03	0,02	0,02	0,02	0,02	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,03	0,01	
	19	0,10	0,10	0,05	0,02	0,02	0,02	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,01	0,01	0,01
	20	0,10	0,10	0,05	0,02	0,02	0,03	0,02	0,01	0,02	0,01	0,01	0,00	0,00	0,01	0,01	0,00	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,01	0,01
	24	0,11	0,10	0,06	0,03	0,03	0,04	0,02	0,02	0,02	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,01	0,01	0,02	0,00	0,00	0,00	0,00	0,00	0,01
	21	0,10	0,10	0,05	0,03	0,03	0,04	0,03	0,02	0,02	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,02	0,02	0,02	0,01	0,00	0,00	0,00	0,00	0,02
	23	0,12	0,12	0,07	0,04	0,05	0,05	0,04	0,03	0,03	0,02	0,02	0,02	0,02	0,03	0,03	0,02	0,02	0,02	0,03	0,01	0,01	0,00	0,00	0,00	0,02
	22	0,14	0,14	0,08	0,04	0,04	0,04	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,03	0,03	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,00
Column Average		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
exp		1,1	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	

On the last page you will find the three diagrams.

Diagram 1 showing **Gini-Simpson index** and **Shannon Equitability**

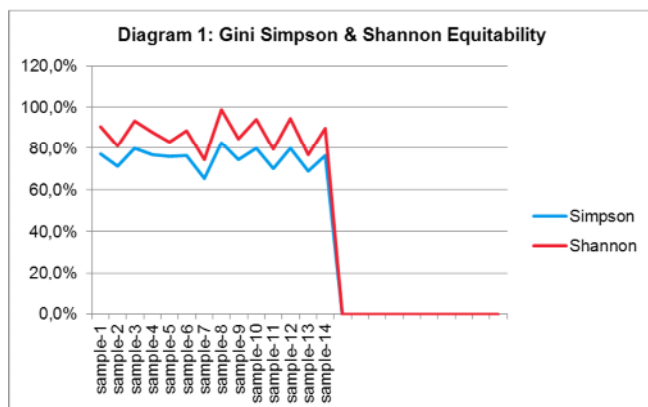


Diagram 2 showing the **average proportional distribution** for all classes/categories:

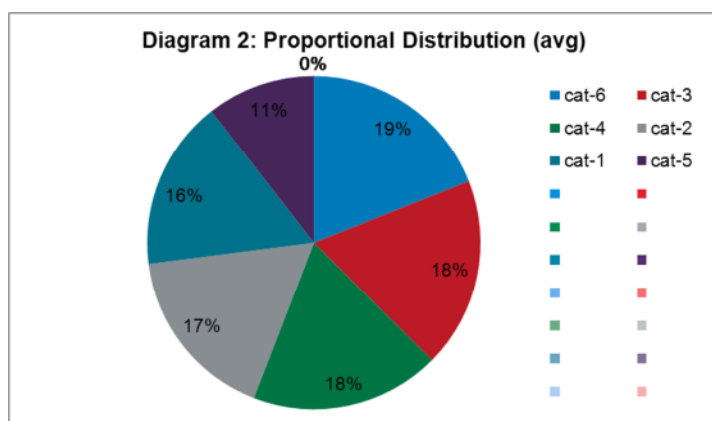
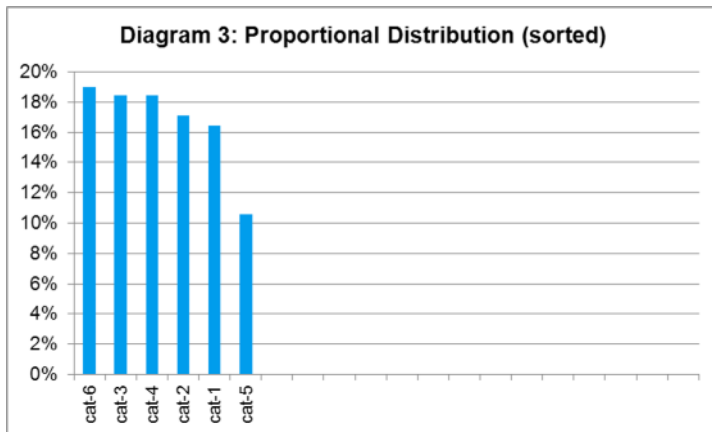


Diagram 3 showing the proportional distribution sorted from largest to smallest proportion (relative abundances)



How to use the template

1. Prepare input data work sheet and table.

Header (has to start always in the first row and column of the input worksheet):

1	Title									1
2	No of Categories	6	Number of classes/categories (2 -20)							
3	No of Samples	14	Number of samples (2 - 24)							
4	Referenz Row	9	Reference row for input data on this worksheet							
5	Reference Column	24	Reference colum for input data on this worksheet							

Input Table (raw input data):

		1	2	3	4	5	6	7	8
Dataset 1		cat-1	cat-2	cat-3	cat-4	cat-5	cat-6	cat-7	cat-8
1	sample-1	0,131	0,131	0,04	0,168	0,159	0,371		
2	sample-2	0,053	0,146	0,256	0,034	0,075	0,437		
3	sample-3	0,181	0,181	0,251	0,251	0,043	0,092		

Proportional abundance (calculated from raw input data):

s-sort	w	0	cat-1	cat-2	cat-3	cat-4	cat-5	cat-6	cat-7	cat-8
1	7%	sample-1	13%	13%	4%	17%	16%	37%	0%	0%
2	7%	sample-2	5%	15%	26%	3%	7%	44%	0%	0%
3	7%	sample-3	18%	18%	25%	25%	4%	9%	0%	0%
4	7%	sample-4	4%	33%	19%	18%	4%	23%	0%	0%

Columns

Column	Column name	Description
1	s-sort	sequence for manual sorting in result sheet
2	w	weight (future extension) –1/N
3	Sample	Sample Name (Text)
4 to N	Cat-1 to n	Rel. abundance of classes/categories

Each sample gives one row in the table. The sum of all relative abundances has to be 100%

You can have several datasets on the input sheet and select them by changing the reference row for input data.

2. Select data sheet (input name of worksheet containing input data table)

Worksheet	input
-----------	-------

Parameter Selections:

AHP

AHP
scale max 9

If your input data are priority distributions based on AHP (Analytic Hierarchy process), you can specify the maximum score of the AHP scale used. Standard is the linear 1-9 scale with scale max =9

Partitioning (Matrix)

Weight: 0	Sort: 4 2	Parameter: 28	Horn
-----------	-----------	---------------	------

Weight (future extension)

- 0: all samples same weight
- 1: weighted according weights given in input sheet

Sort

- 0: unsorted Sequence of samples as given in the input sheet
- 1: sorted by α -diversity From max. to min alpha diversity (left to right)
- 2: sorted by β -diversity From max. to min beta diversity (left to right)
- 3: manual Manual as given by sequence in column s-sort of input sheet
- 4: cluster Clustered: Arranged by consecutive similarity
The second parameter determines the number of the first sample to be displayed in the matrix

Parameter (selects parameter displayed in matrix for all pairs of classes/categories):

- 25 – alpha Shannon α -diversity
- 26 – gamma Shannon γ -diversity
- 27 – beta Shannon β -diversity
- 28 – Horn index
- 29 – AHP consensus index

Clustering

To find clusters of similar data samples use Sort parameter “4” and vary the first sample to be displayed in the matrix, until you see clusters of low beta or high similarity along the matrix diagonal.

Example of clusters with high similarity

	1,00	0,99	0,99	0,99	0,98	0,98
1,00		1,00	1,00	1,00	0,97	0,97
0,99	1,00		1,00	1,00	0,96	0,96
0,99	1,00	1,00		1,00	0,96	0,96
0,99	1,00	1,00	1,00		0,96	0,96
0,98	0,97	0,96	0,96	0,96		1,00
0,98	0,97	0,96	0,96	0,96	1,00	

Please make a reference to the author and website, when using the template in your work:

Goepel, K. D., BPMSG Diversity Calculator, version xx.xx.xx – <http://bpmsg.com>, Singapore 2013

For questions, feedback, suggestions please contact the author under <http://bpmsg.com>

Annex Mathematical relations and formulas used

A. General

1. *Alpha Entropy* for N categories/classes ($N = 2$ to 20) and K samples ($K = 2$ to 24)

$$H_{\alpha} = -w_1 \sum_{i=1}^N p_{i1} \ln p_{i1} - w_2 \sum_{i=1}^N p_{i2} \ln p_{i2} + \dots + w_K \sum_{i=1}^N p_{iK} \ln p_{iK}$$

with p_{ij} relative abundance (frequency, priority, share) of class i and sample j

and w_i statistical weights of samples; $\sum_{i=1}^K w_i = 1$

2. *Gamma Entropy* for K samples

$$H_{\gamma} = \sum_{i=1}^N - (w_1 p_{i1} + w_2 p_{i2} + \dots + w_K p_{iK}) \ln (w_1 p_{i1} + w_2 p_{i2} + \dots + w_K p_{iK})$$

3. *Beta Entropy*

$$H_{\beta} = H_{\gamma} - H_{\alpha}$$

4. *True Diversity Order 1*

$${}^1D_{\alpha,\beta,\gamma} = \exp H_{\alpha,\beta,\gamma}$$

B. Homogeneity measures

5. *MacArthur M*

$$M = 1 / {}^1D_{\beta}$$

6. *Relative Homogeneity 1S for N classes*

$${}^1S = [M - 1/N] / [(1 - 1/N)]$$

7. *AHP consensus S^**

$$S^* = [M - D_{\alpha \min} / D_{\gamma \max}] / [1 - D_{\alpha \min} / D_{\gamma \max}]$$

C. In Table 1

1a *Shannon Equitability*

$$E_H = H_{\alpha} / \ln N$$

1b *Simpson Dominance*

$$\lambda_{k \text{ Simpson}} = \sum_{i=1}^N p_i^2$$

1c *Gini-Simpson Index*

$$\lambda^{-1} = 1 / \lambda_{k \text{ Simpson}} = 1 / \sum_{i=1}^N p_i^2$$

1d *Hill numbers*

$${}^1D = \exp H_{\alpha}$$

$${}^2D = 1 / \lambda_{k \text{ Simpson}}$$

D. Matrix

1. *Alpha Entropy for N categories/classes (N = 2 to 20) and 2 samples k and l*

$$H_{\alpha kl} = -w_k \sum_{i=1}^N p_{ik} \ln p_{ik} - w_l \sum_{i=1}^N p_{il} \ln p_{il}$$

$$H_{\alpha} = \frac{w_k H_{\alpha k} + w_l H_{\alpha l}}{w_k + w_l}$$

For diagonal $k = l$: $H_{\alpha} = H_{\alpha k}$

2. *Gamma Entropy for k categories/classes 2 samples k and l*

$$H_{\gamma kl} = -\sum_{i=1}^N \frac{w_k p_{ik} + w_l p_{il}}{w_k + w_l} \ln \frac{w_k p_{ik} + w_l p_{il}}{w_k + w_l}$$

For diagonal $k = l$

$$H_{\gamma kk} = -\sum_{i=1}^N p_{ik} \ln p_{ik} = H_{\alpha k}$$