

Polyphenols data processing

Initial table

In a first place, we have to know how data is represented. In this case, we have a table, where each row corresponds to an item food with information about their polyphenol content. In the next table, we can see a simple representation of the content that we can find in the database. For our purpose, we don't need all the information in it, so we have crossed out the fields that we don't need.

<i>Food</i>	<i>Food group</i>	<i>Food subgroup</i>	<i>Compound group</i>	<i>Compound subgroup</i>	<i>Compound</i>	<i>Mean</i>	Min	Max	sd	n	N	unit	exp.
F1	A	A1	Polyphenols, total	Polyphenols, total	Polyphenols, total	1.0							
F1	A	A1	G1	S1	C1	1.5							
F1	A	A1	G1	S1	C2	2.1							
F2	A	A2	G1	S2	C3	3.1							
F2	A	A2	G1	S1	C2	0.8							
F2	A	A2	G2	S3	C4	0.5							
F2	A	A2	G2	S3	C5	0.2							
F3	B	B1	G1	S2	C3	1.2							
F3	B	B2	Polyphenols, total	Polyphenols, total	Polyphenols, total	0.9							
F4	B	B3	G2	S4	C1	3.3							

Polyphenols total

The first value we need to get is **Polyphenols, total**. As a fact, is important to know that this value is not calculated for every item food in the table. So, for our purpose, we have to look into each row, and get that ones which have **polyphenols, total** field filled.

This value will be another column of the final database, which is going to be called **polyphenols, total**.

<i>Food</i>	<i>Food group</i>	<i>Food subgroup</i>	<i>polyphenols, total</i>
F1	A	A1	1.0

<i>Food</i>	<i>Food group</i>	<i>Food subgroup</i>	<i>polyphenols, total</i>
F2	A	A2	0.9
F3	B	B1	
F4	B	B3	

Polyphenols B

As we said before, not every item in the database has **Polyphenols, total** filled. That's the reason why is interesting to get that value based on the sum of all the individual polyphenols that we have stored for each item in the database. As we could see in the first table of this document, we have, for each item food, the values of the individual polyphenols. If we sum all this values (we have to exclude **polyphenols total** field for each item), we get the next column that we need to add to our final database. This new column is called **Polyphenols B**. In the next table we can see the calculation for this column.

<i>Food</i>	<i>Food group</i>	<i>Food subgroup</i>	<i>polyphenols B</i>
F1	A	A1	$1.5+2.1 = 3.6$
F2	A	A2	$3.1+0.8+0.5+0.2 = 4.6$
F3	B	B1	1.2
F4	B	B3	3.3

Polyphenols (families)

As well as we have to calculate the total value of polyphenol for every item food of the database, we also need to know the value of different polyphenols families (also for each item food). In this case, for each item, we have to calculate the sum of the individual polyphenol that are part of a same group. That will generate new columns in the final database (as many as families of polyphenols).

In the next table, we can see the result of this processing.

Food	Food group	Food subgroup	G1	G2
F1	A	A1	3.6	0.7
F2	A	A2	3.9	
F3	B	B1	1.2	

Food	Food group	Food subgroup	G1	G2
F4	B	B3		3.3

Polyphenols (subfamilies)

In the previous section, we got new columns from the values of different polyphenol families. We need to repeat this steps, but in this case, we'll considerate subfamilies/subgroup instead of families. In the next table, we can see the result of this processing (similar process to the families values). Again, we'll obtain as many columns as subfamilies of polyphenols.

Food	Food group	Food subgroup	S1	S2	S3	S4
F1	A	A1	3.6			
F2	A	A2	0.8	3.1	0.7	
F3	B	B1		1.2		
F4	B	B3				3.3

Polyphenols (individual)

At last, we have to add, as columns, every individual polyphenol that appears in the table, because, because of it is the base of our calculations. In this way, we'll have, for every item food, all the individual polyphenols stored in database. Notice that, for a lot of items, it can happen that we don't have a value for that measure.

Food	Food group	Food subgroup	C1	C2	C3	C4	C5
F1	A	A1	1.5	2.1			
F2	A	A2		0.8	3.1	0.5	0.2
F3	B	B1			1.2		
F4	B	B3	3.3				

Final

One time we have all the new columns, we have to concatenate them, in order to get a new table where: - Every row corresponds to an unique item food - Columns have the values of the different polyphenol calculations (based on the stored values we have for each item).

In the next table, we can see the result for this example.

Food	Food group	Food subgroup	polyphenols, total	polyphenols B	G1	G2	S1	S2	S3	S4	C1	C2	C3	C4	C5
F1	A	A1	1.0	3.6	3.6		3.6				1.5	2.1			
F2	A	A2		4.6	3.9	0.7	0.8	3.1	0.7			0.8	3.1	0.5	0.2
F3	B	B1	0.9	1.2	1.2			1.2					1.2		
F4	B	B3		3.3		3.3				3.3	3.3				