

Calculating a Divergence Rate

Calculating a divergence rate is a key step in quality control for coding. The rate of divergence is the percentage of responses on which the researchers diverged when redundantly coding their study. A high rate of divergence may indicate that the questions or answers need to be modified, or that one of the researchers requires additional training in coding. A high rate of divergence also signals to the supervisor that more jurisdictions should be redundantly coded, until the rate of divergence is reduced.

A low rate of divergence may indicate that the questions and answers are clear, the coding scheme is working, and that the coders do not require additional training. However a low rate of divergence may also be a coincidence, if some jurisdictions are easier to code than others, or if the coders are experts on the topic of study. Therefore, it is important to take outside factors into account before reducing the number of records that need to be redundantly coded.

Although more complex projects can have a higher rate of divergence, any rate of divergence over 10% indicates that some measures should be taken to reduce it, and that more records should be redundantly coded. A rate of divergence under 5% is low and indicates that fewer records need to be redundantly coded.

It is normal for the rate of divergence to be higher at the beginning of a project then gradually lessen, because in the beginning of the project few coding decisions have been made. Coding decisions may include defining terms, clearing up questions, or removing answer choices that do not work. In sum, as the project moves along, coding decisions are made to clarify the coding scheme, thus the rate of divergence should steadily decline over time.

Microsoft Excel can be used to calculate a rate of divergence after coding. Below you will find a step by step guide on calculating the divergence rate.

Calculating the Rate of Divergence

An extract of the coding data for your study is necessary to calculate a rate of divergence, and should include a column for each variable in the study. Each answer choice becomes a variable. Coding answers should then be organized by researcher and by jurisdiction, to indicate whether each coder selected an answer choice or not. Below is an example of what this should look like in Excel, using an outpatient commitment study.

State	oc-coder	oc-iteration	oc-criteria_Dangerous to self	oc-criteria_Dangerous to others
AZ	MA	1	0	0
AZ	BK	0	1	1
CA	BK	1	0	0
CA	MA	0	0	0

The first column is the jurisdiction, the second column is the coder initials, the third column is the iteration of the law, and the following columns are potential answer choices. A 0 indicates that the answer choice was not selected, whereas a 1 indicates that it was (Except for the iteration column, which indicates what version of the law is being coded in a longitudinal dataset. In this instance, there is an iteration 0 as well to indicate that a record was redundantly coded). Finally, there is a blank row between each jurisdiction, which leaves space for the formulas which will be used.

The next step after setting up your data is to identify on which answer choice the researchers diverged. This can be done by using a formula in the blank cell under an answer choice, which can be copy pasted for all answers and jurisdictions. The formula is as follows:

=if (click first answer choice=second answer choice,0,1) press enter.

Below is an example of what this should look like in excel:

	A	B	C	D	E
1	State	oc-coder	oc-iteration	oc-criteria_Dangerous to self	oc-criteria_Dangerous to others
2	AZ	MA	1	0	0
3	AZ	BK	0	1	1
4				1	1
5	CA	BK	1	0	0
6	CA	MA	0	0	0
7				0	0

As you can see, where it says “click first answer choice”, the supervisor clicks on the top answer choice for AZ in column D, then =, and then clicks on the bottom answer choice for AZ in column D. The resulting formula is =IF(D2=D3,0,1). This formula, which appears in cell D4, can then be copied and pasted in every blank cell under the answer choices for each jurisdiction, and the values will automatically change to compare the correct answer choices.

This formula will put a 0 under answer choices with no divergence, and a 1 under answer choices which did diverge. Below is an example of what this should look like in excel, continuing with the outpatient commitment study:

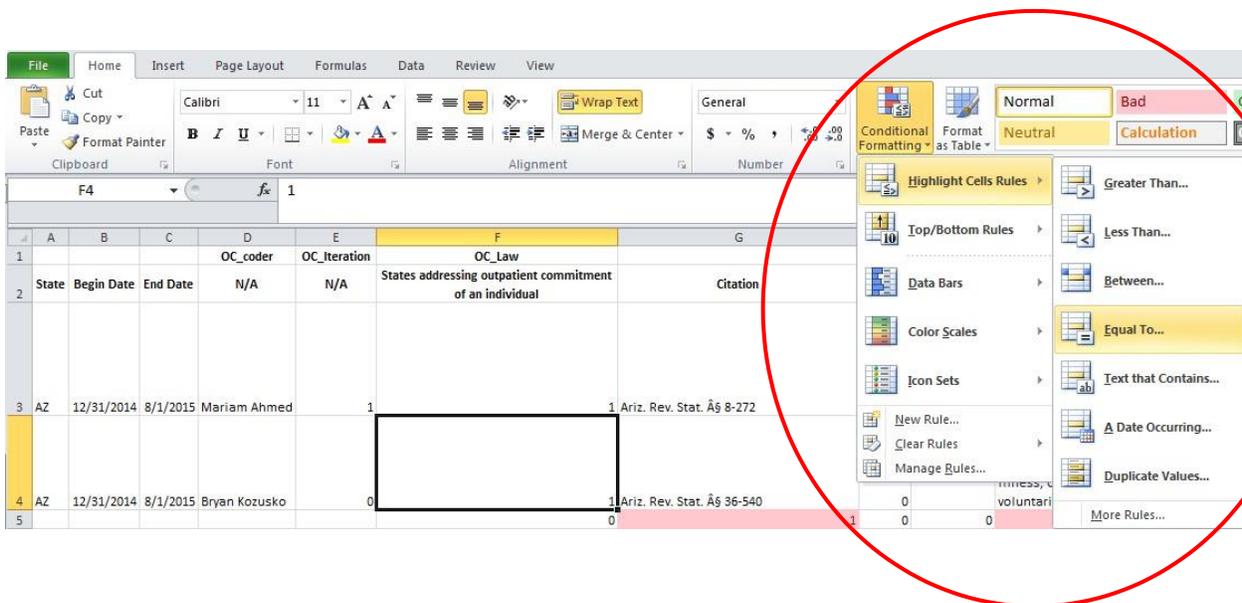
State	oc-coder	oc-iteration	oc-criteria_Dangerous to self	oc-criteria_Dangerous to others
AZ	MA	1	0	0
AZ	BK	0	1	1

			1	1
CA	BK	1	0	0
CA	MA	0	0	0
			0	0

For AZ (Arizona), the first two answer choices diverged, so there is a 1 in each blank cell. However, for CA, the first two answer choices were the same, so there is a 0 in each blank cell

The next step is to highlight divergences so that they can be more visible on the spreadsheet. This is an optional step, which simply makes it easier to read divergences in the spreadsheet. This is done by highlighting the cells which were used for the divergence formula, and going into conditional formatting on the home tab of excel. It is important to only highlight the cells used for the formula, because otherwise this step will also highlight other “1”s in the spreadsheet.

To highlight divergences, click on “highlight cells rules”, “equal to”, and type 1 in the left side text box which says, “Format cells that are EQUAL TO”. In the right hand drop down menu, you can select the color which will be used to highlight divergences. Below is an example of what this should look like in excel:



This is what your cells should look like after you have highlighted divergences:

State	oc-coder	oc-iteration	oc-criteria_Dangerous to self	oc-criteria_Dangerous to others
AZ	MA	1	0	0
AZ	BK	0	1	1

			1	1
CA	BK	1	0	0
CA	MA	0	0	0
			0	0

This step is also useful for the supervisor to be able to quickly identify which questions were causing divergences, so that they can be discussed in a meeting with the researchers to reduce the rate of divergence for ensuing jurisdictions.

The next step is to add up the total number of divergences in each jurisdiction. The summation formula should be used for each jurisdiction, and should go in the last column of the spreadsheet, after the last answer choice, in the divergence formula row. The formula is as follows:

=sum(click on first entry equation:click last variable entry)

Below is an example of what this should look like in excel:

BG4		fx =SUM(D4:BF4)				
	A	BC	BD	BE	BF	BG
1	State	oc-discharge	oc-fapossession	oc-farelief	oc-fareporting	
2	AZ	0	1	1	1	
3	AZ	0	1	1	1	
4		0	0	0	0	9
5	CA	0	1	1	0	
6	CA	0	1	1	1	
7		0	0	0	1	9

As you can see, where it says “click on first entry equation”, the supervisor clicked on cell D4, which contains the first divergence calculation formula. Then they typed :, and clicked on the last cell which contains a divergence formula for this jurisdiction, cell BF4. This yields the total number of divergences for this jurisdiction, 9 in this case.

The next step is to add up the total number of answer choices in the study, and to multiply that number by the number of jurisdictions being evaluated in this divergence calculation. The total here will represent the total number of potential divergences (answer choices X jurisdictions). To quickly count the number of answer choices, simply highlight the top row of your spreadsheet starting with the first answer choice and ending with the last, and at the bottom right corner of excel it should say “Count: x” to tell you how many answer choices you have. This is an example of what this should look like in excel:

	A	AW	AX	AY	AZ	BA	BB	BC	BD	BE
1	State	oc-durationmax	oc-durationmod	oc-modcond Physician evaluation	oc-modcond Court order	oc-modcond Prior convictions	oc-modcond Prior commitment	oc-discharge	oc-fapossession	oc-fareleief
2	AZ	3	1	1	0	0	0	0	1	1
3	AZ	3	1	1	1	0	0	0	1	1
4	AZ	0	0	0	1	0	0	0	0	0
5	CA	2	1	1	1	0	0	0	1	1
6	CA	2	1	1	1	0	0	0	1	1
7	CA	0	0	0	0	0	0	0	0	0
8	DE	1	1	1	0	0	0	0	1	1
9	DE	1	1	1	1	0	0	0	0	0
10	DE	0	0	0	1	0	0	0	1	1
11	MS	1	1	1	0	0	0	0	0	0
12	MS	1	1	1	0	0	0	0	1	1
13	MS	0	0	0	0	0	0	0	1	1
14	MT	1	1	1	1	1	0	0	0	0
15	MT	1	1	0	1	0	1	0	0	0
16	MT	0	0	1	0	1	1	0	0	0
17	NJ	2	1	1	1	0	0	1	1	1
18	NJ	1	1	1	1	0	0	1	0	0
19	NJ	1	0	0	0	0	0	0	1	1
20	NY	3	1	1	1	0	0	0	1	1
21	NY	3	1	0	1	0	0	0	1	1
22	NY	0	0	1	0	0	0	0	0	0
23	TX	1	1	1	1	1	0	0	1	1
24	TX	1	1	1	0	1	0	1	1	1
25	TX	0	0	0	1	0	0	1	0	0
26	UT	2	1	1	0	0	0	0	1	1
27	UT	2	1	1	1	0	0	0	1	1
28	UT	0	0	0	1	0	0	0	0	0
29	VT	1	1	0	1	0	0	0	0	0
30	VT	1	1	0	1	0	0	0	0	0
31	VT	0	0	0	0	0	0	0	0	0
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										
47										
48										



In our outpatient commitment study, we have 55 total answer choices. We are looking at two jurisdictions in this example, Arizona and California, so we would multiply 55 by 2 (number of jurisdictions being compared) for a total of 110 total variables.

The final step is to divide the total number of divergences by the total number of variables. In this instance, we had 18 total divergences, and 110 total variables. 18 divided by 110 is 0.1636, which is equivalent to a divergence rate of 16.36%.