

Calculating LongString in Excel to Detect Careless Responders

Richard N. Landers¹

¹Affiliation not available

November 20, 2020

Careless responding is one of the most fundamental challenges of survey research. We need our respondents to respond honestly and with effort, but when they don't, we need to be able to detect and remove them from our datasets. A few years ago, Meade and Craig (2012) published an article in *Psychological Methods* exploring a significant number of techniques for doing exactly this, ultimately recommending a combination of three detection techniques for rigorous data cleaning, which, let's face it, is a necessary step when analyzing any internet survey. These techniques are even-odd consistency, maximum longstring, and Mahalanobis D:

1. **The even-odd consistency index** involves calculating the subscale means for each measure on your survey, split by even and odd items. For example, the mean of items 1, 3, 5, and 7 would become one subscale whereas the mean of items 2, 4, 6, and 8 would become the other. Next, you take all of the even subscales and pair them with all of the odd subscales across all of the measures in your survey, calculate a correlation, and then apply the Spearman-Brown prophecy formula to adjust the value up to a scale of -1 to 1.
2. **Maximum LongString** is the largest value for LongString across all scales on your survey, where LongString is the number of identical response in a row. Meade and Craig recommended LongString would be most useful when the items were randomly ordered.
3. **Mahalanobis D** is calculated from the regression of scale means onto all the scores that inform them. In a sense, you are trying to see if responses to individual items correspond with the scale mean they created consistently across individuals. Some conceptualizations of this index regress participant number onto scores, which conceptually accomplishes basically the same thing.

In all three cases, the next step is to create a histogram of the values and see if you see any outliers.

Calculating Careless Responding Indices

Of these three, Mahalanobis D is the most easily calculated, because saving Mahalanobis D values is a core feature in regression toolkits. It is done easily in SPSS, SAS, R, etc.

The second, the even-odd consistency index, is a bit harder but still fundamentally not too tricky; you just need to really understand how your statistical software works. Each step, individually, is simple: calculate scale means, calculate a correlation, apply a formula.

The third, Max LongString, is the most intuitively understandable but also, often unexpectedly, the most difficult to calculate. I imagine that the non-technically-inclined generally count by hand – “this person has a maximum of 5 identical answers in a row, the next person has 3...”

[An SPSS macro already exists to do this](#), although it's not terribly intuitive. You need to manually change pieces of the code in order to customize the function to your own data.

Given that, I decided to port the SPSS macro into Excel and make it a little easier to use.

An Excel Macro to Calculate LongString

```
Function LongString(cells As Range)      Dim cell As Range      Dim run As Integer      Dim firstrow As Boolean      Dim maxrun As Integer      Dim lastvalue As String      firstrow = True      run = 1      maxrun = 1      For Each cell In cells      If firstrow = True Then      firstrow = False      lastvalue = cell.Value      Else      If cell.Value = lastvalue Then      run = run + 1      maxrun = Application.Max(run, maxrun)      Else      run = 1      End If      lastvalue = cell.Value      End If      Next cell      LongString = maxrun      End Function
```

To Use This Code Yourself

1. With Excel open, press Alt+F11 to open the VBA Editor.
2. Copy/paste the code block above into the VBA Editor.
3. Close the VBA Editor (return to Excel).
4. In an empty cell, simply type `=LONGSTRING()` and put the cell range of your scale's survey items inside. For example, if your first scale was between B2 and G2, you'd use `=LONGSTRING(B2:G2)`
5. Repeat this for each scale you've used. For example, if you measured five personality dimensions, you'd have five longstrings calculated.
6. Finally, in a new cell use the `=MAX()` function to determine the largest of that set. For example, if you put your five LongStrings in H2 to L2, you'd use `=MAX(H2:L2)`

That's it! Importantly, the cells needs to be in Excel in the order they were administered. If you used randomly ordered items, this adds an additional layer of complexity, because you'll need to recreate the original order for each participant first before you can apply LongString. That takes a bit of Excel magic, but if you need to do this, I recommend you read up on `=INDIRECT()` and `=OFFSET()`, which will help you get that original order back, assuming you saved that item order somewhere.

Final Steps

Once you have Max LongString calculated for each participant, [create a histogram](#) of those values to see if any strange outliers appear. If you see clear outliers (i.e., a cluster of very high Max LongString values off by itself, away from the main distribution), congratulations, because it's obvious which cases you should drop. If you don't see clear outliers, then it's probably safer to ignore LongString for this analysis.

Another potential issue with Max LongString is that if you have scales with varying numbers of items, shorter scales can be lost when you calculate the Max. To avoid that problem, try converting each of your longstrings into a proportion. For example, for a 4-item scale, try `=LONGSTRING(H4:H7)/4` and then calculate MAX() on those results.

References

Meade AW, & Craig SB (2012). Identifying careless responses in survey data. *Psychological Methods*, 17 (3), 437-55.