EPBI 414 (Fall 2016) - Assignment 2

Simple Forms and Data Concepts

Overview

This homework consists of three parts. For the first part, you'll answer some questions about the material we've covered for the last few weeks. Next, you'll review data and see if it complies with an existing data dictionary. Finally, you'll annotate another CRF and create a data dictionary from it.

Logistics

Submit your assignment in a .zip file labeled in the following matter:

<Case ID>_EPBI414_Fall2016_A2.zip

So, if your Case ID is tar9, you would submit the following zip file:

tar9_EPBI414_Fall2016_A2.zip

Part 1

Submit a written document with your answers to the following questions.

1. Some epistemic models that were discussed in lecture included quantification, classification, and networking. Come up with three examples of data fitting into each of these models. Then, come up with another epistemic model that might be useful to organize data. [12 points]

2. Come up with an overly broad research question. Then, develop three *ante hoc* abstractions of that question that would work better in a case report form or other quantitative study. [8 points]

3. Write a simple categorical question where you might not be able to tell the difference between data that is actually missing, and data that is missing due to data entry errors. Then, explain how you would rewrite the question and answers to eliminate the issue. [5 points]

4. Create a numeric array with two dimensions, one of size 5 and one of size 3. Fill the array with random integers of your choice. Then, show how you could represent the same array as a one-dimensional array of size 3, where each element is another one-dimensional array of size 5. [5 points]

Part 2

Review the ODS file titled "form04_sample_data.ods". ODS is the file extension for an OpenDocument Spreadsheet. This file should be compatible with Microsoft Excel as well as alternative spreadsheet programs like LibreOffice Calc. Normally, I use CSV (comma-separated values) files for most things, but Excel and other spreadsheet software have a nasty habit of mangling the data in them when you open them.

This file contains a spreadsheet, "Raw Data", with 15 records of data produced by the CRF and data dictionary reviewed in class. However, there are several errors throughout the spreadsheet - places where the data does not conform to the data dictionary. Review the data and list each instance, specifying why the data does not meet the requirements in the data dictionary. Include this with your written answers from Part 1. [20 points]

Part 3

Review the Demographics CRF included with the assignment. Using a method of your choice, annotate this CRF with the information necessary to document the data structure. Be sure to remember that all the fields on the form need to be annotated, and that each field should have an appropriate type for the data it collects. Hint: you should find a total of 20 fields on the form. [40 points]

Next, use this annotated version of the CRF to produce a data dictionary, similar in style to the one used as an example in class. Be sure to include the same information as the example dictionary. The questions on this form are a little more complicated than those in the first example. Be sure to think about what logic you need to apply so that the data collected is complete. Hint: when you have fields that are filled out only under certain circumstances, it's important to document what happens when those circumstances ***are not*** met. [40 points]