Simulating Dichotomous Items

The simplest general strategy for dichotomous items is to simulate continuous variables and then transform them into binary variables coded 0 or 1. This strategy has the advantage of being simple and transparent. If more control is needed, one can instead adopt a strategy based on IRT type models. However, the simpler strategy should suffice for the current project.

Once you simulate your continuous variables, there are two basic strategies for creating binary variables. The first involves computing a new set of variables from the continuous variables. The second strategy involves changing the way that the variables are coded. Both work equally well. Let's take them in order.

One can break down the first strategy into steps. First, as a means of defensive programming, compute the new variables as missing. Second, replace the missing value with 0 for all cases below a specified cut point. Second, replace the missing value with 1 for all cases at or above the specified cut point. Below, I adopt a cut score of 5 just for illustration.

Here is pseudocode in SPSS syntax for continuous variable I1.

```
compute I1Bin = -9.
recode I1Bin (-9 = sysmis).
if (I1 < 5) I1Bin = 0.
if (I1 GE 5) I1Bin = 1.
* Good idea to check your work:
crosstabls /tables= i1 by i1Bin.
```

Here is pseudocode in R for continuous variable I1.

```
I1Bin <- rep(NA, len=length(I1))
I1Bin[I1 < 5] <- 0
I1Bin[I1 >= 5] <- 1
# Good idea to check your work:
table(I1, I1Bin)
```

Here it is in Excel.

I1	I1Bin
	=if(a2 < 5, 0, 1)

The second strategy is to modify the coding of a variable. However, as defensive programming, it is always safest to work with a copy rather than risk messing up your original variable. Here is an SPSS version.

```
recode i1
     (5 thru highest = 1) (else = 0)
     into i1r .
execute.
* Good idea to check your work:
crosstabls /tables= i1 by i1r.
```

Here it is in R.

```
i1r <- cut(i1, breaks=c(-Inf, 5, Inf), right=FALSE)
# Good idea to check your work:
table(i1, i1r)
```

You would essentially use the same code either way in Excel.

Either way, you need to consider two further alternative strategies for determining the difficulty of your items. Two things in the code will affect this. The first is the intercept values that you choose to fix the mean of the continuous variables. The second is the value that you choose for the cut score in the above code. Here is the sample SPSS code that controls the intercepts.

```
* The next set of lines compute the intercepts.
compute I1 = 5.
```

If you mess around with both the intercepts and the cut score, you might have trouble dialing up the exact item difficulties that you want to obtain. A better strategy would be to hold one constant, it does not matter which, and manipulate the other. For example, give items different intercept and use a fixed cut score for all items. Alternatively, hold the intercept fixed for all items, and vary the cut score to produce items with different levels of difficulty.