Chapter 11

SAMPLING

1. Sampling

Unless an audit client is very small almost all auditing relies on sampling rather than looking at all transactions and documents. This is because there simply isn't time to look at all documents and it wouldn't be economically viable to do so. If, however, valid statistical conclusions are to be drawn about a population by relying on a sample, then the sample must be free of bias. In other words every document or transaction of the population has an equal chance of being included in the sample. This is known as statistical sampling.

Sampling methods include:

- **Random sampling.** The best way to remove bias and to obtain statistical sampling is to adopt what's called random selection. Let's say we wanted to look at purchase invoices throughout the year. There might be 20,000 purchase invoices and we want to inspect 20 of them. What you would do is to number the 20,000 invoices consecutively and then use a random number generator to produce 20 numbers and you would then go and look at the corresponding invoices. The difficulty with this approach is that very often the population is not pre-numbered and to set out initially numbering all 20,000 invoices is very time-consuming.
- **Systematic sampling.** As an approximation to pure random selection, systematic selection or systematic sampling might be used. Again, if with 20,000 invoices and wanted to look at about 20 invoices out of that you could do that by looking at about every 1,000th invoice. So what you would do is that near the beginning of the population you would choose an invoice at random and then count through selecting every 1,000th one. Provided there isn't some weird correspondence of every 1,000 invoice being from exactly the same supplier, you are going to get pretty close to random selection.
- **Haphazard sampling** is frequently used because it is rather convenient. It's essentially the auditor opening a file at random and picking the invoice at which the file is opened. There can be obvious problems with this. For example, the file might always open at a slightly thicker invoice or a slightly larger invoice and that invoice could be from the same small group of suppliers. There might be a relatively small chance of the physically small invoice being chosen. It gets even worse because it's open to some abuse. Inevitably if the auditors are closely involved with choosing the invoice and they look at the invoice before making a final choice, and in looking at the invoice may be tempted to choose only those invoices which appear to be correct or which are simple to deal with. This is introducing very obvious bias and reducing the chances of finding transactions or documents which need to be investigated further.
- **Sequence or block selection** will mean choosing 20 invoices all in a sequence. They may all be from different suppliers, but they are probably all from the same week or even the same day and to be no assurance that the controls which operated during that restricted period have operated during the whole financial year.
- **Stratification**. If we know that there are 20,000 invoices, 10 of those are above 100,000 then it might make sense to make sure we choose at least all of those 10 invoices plus

another 10 chosen randomly. Stratification means dividing your population into different layers, usually based on size, and making sure your sample is deliberately biased so that the larger and more significant transactions have a greater chance of being covered. This does rank as statistical sampling provided samples are drawn from every layer.

- **Monetary unit sampling**. This is rather more complex…

2. Monetary unit sampling

Monetary unit sampling can be regarded as a form of stratified sampling. Here is an example to indicate

Invoice value ($)	Cumulative invoice value		
80	80		
70	150		5,000/4 = 1,250
400	550		
90	640	←	Choose first at random – say, 605
1,600	2,240	←	Then: 1,855 (= 605 + 1,250)
20	2,260		
700	2,960		
50	3,100		
1,010	4,020	←	Then: 3,105 (= 1,855 + 1,250)
80	4,100		
30	4,130		
600	4,730	←	Then: 4,355 (= 3,105 + 1,250)
380	5,110		

What we have is a list of say customer invoices 80, 70, 400, 90, all the way down to 380.

The right hand column of the table is a cumulative total, so the first one is 80, then 80 plus 70 is 150, 150 plus 400 is 550, 550 plus 90 is 640, so our total receivables is 5110.

We want to look at four invoices out of these receivables. So you take the total, and if we round it to 5000 and divide by 4 that give 1250. Choose the first interval at random, here is say 605, and then go up 1250 at a time. So after 605 plus 1250 will be 1,855, plus 1250 will be 3,105, plus 1250 will be 4,355 and you see where a cumulative total of those values lie. So 1,855 falls within the cumulative total of 2240 and that is opposite the account of value 1600. The next one 3,105 falls within the cumulative 4020 and that points to account with value 1010.

What this process does is to give you a better chance at picking the higher value balance. For example the balance of $1600 spans a range of $1600 and there is a much higher chance of that balance being chosen. The next balance of $20 only spans a value of $20, and has little chance of being picked. So monetary unit sampling skews the selection towards picking larger values and therefore more significant items for you to check.