

1. Review data from previous 2021 study:
 - a. Calculate the account class mean and variance for the “sampled” strata of the previous study to use as the proxy mean and variance for the current study.
 - b. Calculate Fisher’s minimum sample size required for application of the normal approximation to calculate confidence limits for positively skewed distributions.
 - c. Export summarized data to spreadsheet.
 - d. Inspect for breakpoint from previous study.
 - e. Determine normality or skewness of data.
2. Group current study data by document number:
 - a. Import current study population data from spreadsheet to database.
 - b. Group current study invoices by document number.
 - c. Verify that the total invoice amount matches the total amount from the spreadsheet.
 - d. Export to spreadsheet and again verify that the total invoice amount matches the total amount from the spreadsheet. Add cumulative sum column.
3. Determine sample size using mean and variance of the “sampled” strata from the previous 2021 study and number of census records and weight from current study. Sample sizes are calculated to yield a 90% confidence interval with less than 10% sampling error. Use a minimum of 30 sample points or Fisher’s minimum sample size for positively skewed distributions and add additional records to sample to account for missing data and unknown differences between the previous and current study’s data.
4. Setup for sample selection:
 - a. In the population spreadsheet, calculate a running total based on the absolute value of the invoice amount for each sample record sorted in descending order. Sample records will be selected if a random number falls between these two values. This applies the probability proportional to size with replacement sampling so that larger sample records have a higher probability of being selected. The absolute value of the invoice amount is used so that negative values have the same probability as positive values of being selected.
 - b. Export to database.
 - c. Verify that the total invoice amount matches the total amount from the original population.
5. Generate random numbers:
 - a. Generate a list of random numbers using the function “RAND()” multiplied by the total of the absolute value of the sample records.
 - b. Copy random values and paste the values only to a new column (this is due to the fact that Excel will re-calculate the random numbers every time any calculation is performed and we need them to be static).
 - c. Copy and paste the random numbers to the database.
6. Select sample records:
 - a. Select the records for each account class if a random number falls between the sample record cumulative min and cumulative max.
 - b. Append census records to selected sample records.
7. Send invoices selected to Revenue Analysis for service period lookup.

Sampling Calculations

Design sample for 90% confidence of 10% precision (90/10) for each account class split into a census stratum and a sample stratum using the following equations:

$$n_{k,s} = (1 - W_{k,census})^2 \frac{Z^2 S_{k,2020}^2}{d^2 - \bar{x}_{k,2020}^2} * 1.5$$

where

$n_{k,s}$ = sample size for sample stratum only of account class k

$W_{k,census}$ = census stratum expense stated as a percent of total account class k expense dollars

Z = 1.645 (90% confidence based on normal distribution)

$S_{k,2020}$ = standard deviation of 2020 lead day mean for account class k

d = precision required (10%)

$\bar{X}_{k,2020}$ = 2020 estimated lead day mean for account class k

A minimum of 30 sample points or the sample size calculated by Fisher's adjustment to account for a positively skewed distribution is required.

If

$n_{k,c}$ = census stratum size (top $W_{k,census}$ of invoices)

Then

$$n_{k,total} = n_{k,c} + n_{k,s}$$

Randomization Method

Random numbers were generated:

RAND() * Total Dollar Amount of Sample Invoices

so that probability proportional to size (pps) sampling could be employed. Sample invoices were sorted in descending dollar amount order. If a random number fell within the cumulative amount of the invoice, that invoice was chosen as a sample point.