**Statistically Valid Sampling**

Sampling is an essential step in determining the quality of work that is being delivered without having to check all the reports or jobs delivered. Good quality sampling is characterized by the sampling technique used and the sample size picked. Obtaining a sample that is appropriate in both regards is critical to having a good understanding of the quality of work delivered. Sampling must be done at a job level since accuracies are measured at a job level.

**Sampling Technique**

Using a random sampling technique will ensure all jobs delivered have an equal opportunity to be picked as a sample. Random sampling is both easy to use and can give an accurate representation of all jobs delivered.

**Sample Size Determination**

Determining the correct sample size will help us get an accurate measure of the quality of work while using the resources required for sampling in an optimal manner. Using a larger sample size could provide us better accuracy in determining the population (population is the set of all jobs delivered); however, this will exhaust more resources for sampling. On the flip side, a smaller sample size would help conserve resources but may not necessarily provide a good understanding of the quality of the population. Also, a smaller sample size is susceptible to higher variation in the population (e.g., if the population contains jobs with accuracies that are highly varied). Calculating the minimum sample size helps us determine the correct sample size required.

**Minimum Sample Size Calculation**

Minimum sample size (MSS) is calculated using this formula:

**Abbreviations Quick Reference**

|  |  |
| --- | --- |
| MSS | minimum sample size |
| SD | standard deviation |
| CI | confidence intervals *or* confidence level  |
| P | precision |

 **MSS = ((CI \* SD)/P)2**

SD refers to the estimated standard deviation of the population. Standard deviation is a measure of variation. A low standard deviation indicates a low variation. If the standard deviation is low then MSS will also be low; the converse is also true.

CI refers to the confidence intervals (also known as confidence level), which determines the probability that the sample will represent the population. This should be set at 95%. Increasing the CI to greater than 95% will increase the probability that the sample represents the population, but this will also increase the minimum sample size required; the converse of the statement is also true, i.e., reducing the CI will result in reduction of MSS.

P refers to precision, which means the accuracy level, in decimal points, that we would want when determining the quality of the population. It is recommended that precision be set at 0.025 for healthcare documentation jobs. What this means, essentially, is if the actual quality of the population is 99.50, the sample will give us an accuracy estimate that will fall between 99.475 and 99.525.

The spreadsheet embedded below provides a calculator for minimum sample size. Once the user inputs the standard deviation, confidence level, and precision, the minimum sample size required is provided.

 

**Data Requirements to Determine Minimum Sample Size**

To calculate the minimum sample size, it is necessary to get the estimated standard deviation of the population (SD). To do this, collect the job level accuracies for all the jobs audited. Using the latest three (3) months of data is ideal. Data should be collected only for unbiased samples. Biased samples like focused audits, special audits, version audits, etc., which concentrate on specific authors, employees, or other specific variables should be excluded from this data set. Standard deviation can be easily calculated in Microsoft Excel using the formula “STDEV.”

**Caveats**

* MSS calculation should not be generalized; i.e., MSS calculated for a particular data set or population should not be used for its subsets because each subset can have a different standard deviation.
	+ ***Example:*** If minimum sample size is calculated for a particular facility, then it should be used only for that facility and should not be used for sampling a specific author, healthcare documentation specialist, or department within that facility as the standard deviation can be different for each.
* It is possible that in case of smaller facilities (or data sets) the minimum sample size calculated would exceed the total number of jobs delivered (or total data points), or the minimum sample size may be too large for a department or facility to audit. Employ subjective decision-making in such cases, keeping in mind that the confidence level of such samples will be less than 95%.
* Each organization or MTSO will have to determine how they will use the resulting numbers in performing their retrospective QA reviews, keeping in mind their budgetary constraints.
* The recommended number of reports to review may be spread over a period of time that aligns with your QA staffing capabilities and budget constraints.
* **Note regarding clinician QA:** If no score is assigned to clinician reviews, these sampling guidelines cannot be used.

**Examples for Calculation of Minimum Sample Size by Facility/Client**

Consider this scenario: We have two facilities (or MTSO clients) for which we need to determine the minimum sample size, and we assume the information below is provided data of the job level accuracy of all audited jobs for the last 3 months. Calculate the standard deviation in Microsoft Excel using the formula “STDEV.” Our example data set is provided below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  | **SD** |
| **Facility1** | 99.89 | 99.56 | 99.76 | 99.87 | 100 | 100 | 99.87 | 99.89 | 99.6 | 100 | 99.84 | **0.1503** |
| **Facility2** | 99.95 | 99.61 | 100 | 99.45 | 100 | 99.95 | 99.87 | 99.32 | 100 | 100 | 99.82 | **0.2443** |

As given in the recommendations, the confidence level is set at 95% and precision is set at 0.025.

**Facility 1:** The standard deviation for Facility 1 is 0.1503. Using the calculator, we get the minimum sample size for Facility 1 as 139 jobs.

|  |  |  |
| --- | --- | --- |
| **Continuous Data** | **Inputs** | **Answer** |
| Standard Deviation |               0.1503  |   |
| Confidence Level (e.g. 95%) | 95.0% |   |
| Precision (e.g., ± 2 units) |                 0.025  |    |
| Minimum Sample Size |   | 139 |

This means for Facility 1 we need to sample at least 139 jobs to have a 95% confidence level that the sample picked is a good representation of all jobs delivered.

**Facility 2:** The standard deviation for Facility 2 is 0.2443. Using the calculator, we get the minimum sample size for Facility 2 as 367 jobs.

|  |  |  |
| --- | --- | --- |
| **Continuous Data** | **Inputs** | **Answer** |
| Standard Deviation |               0.2443  |   |
| Confidence Level (e.g. 95%) | 95.0% |   |
| Precision (e.g., ± 2 units) |                 0.025  |    |
| Minimum Sample Size |   | 367 |

This means for Facility 2 we will need to sample at least 367 jobs to have a 95% confidence level that the sample picked is a good representation of all jobs delivered. Here, a higher number of jobs need to be sampled for Facility 2 as the standard deviation for Facility 2 is high. The standard deviation for Facility 2 is high because job level variation in accuracies is higher for Facility 2 (varies between 99.32 to 100).

**Examples on Calculation of Minimum Sample Size by Healthcare Documentation Specialist**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Job 1** | **Job 2** | **Job 3** | **Job 4** | **Job 5** | **Job 6** | **Job 7** | **Job 8** | **Job 9** | **Job 10** | **SD** |
| **Employee 1** | 100 | 99.75 | 99.5 | 99.75 | 99.75 | 99.75 | 100 | 100 | 99.75 | 99.5 | **0.1845** |
| **Employee 2** | 98.5 | 98.25 | 99 | 98.75 | 98.25 | 98.5 | 99 | 98.25 | 97 | 98.25 | **0.5683** |

As given in the recommendations, the confidence level is set at 95% and precision is set at 0.025.

**Employee 1:** The standard deviation for Employee 1 is 0.1845. Using the calculator, we get the minimum sample size for Employee 1 as 209 jobs.

|  |  |  |
| --- | --- | --- |
| **Continuous Data** | **Inputs** | **Answer** |
| Standard Deviation |               0.1845  |   |
| Confidence Level (e.g. 95%) | 95.0% |   |
| Precision (e.g., ± 2 units) |                 0.025  |    |
| Minimum Sample Size |   | 209 |

This means for Employee 1 we need to sample at least 209 jobs to have a 95% confidence level that the sample picked is a good representation of all jobs delivered.

**Employee 2:**  The standard deviation for Employee 2 is 0.5683.  Using the calculator, we get the minimum sample size for Employee 2 as 1985 jobs.

|  |  |  |
| --- | --- | --- |
| **Continuous Data** | **Inputs** | **Answer** |
| Standard Deviation | 0.5683 |   |
| Confidence Level (e.g. 95%) | 95.0% |   |
| Precision (e.g., ± 2 units) |                 0.025  |    |
| Minimum Sample Size |   | 1985 |

This means for Employee 2 we will need to sample at least 1985 jobs to have a 95% confidence level that the sample picked is a good representation of all jobs delivered. Here, a higher number of jobs need to be sampled for Employee 2 as the standard deviation for Employee 2 is much higher compared to Employee 1. The standard deviation for Employee 2 is high because job level variation in accuracies is higher for Employee 2 (varies between 97 to 99).

**NOTE:** If the sample size is very high and impractical, the confidence level could be lowered to achieve a more practical sample size. In the Employee 2 example, if the confidence level were changed to 80%, the minimum sample size would be reduced to 849.

**Glossary Quick Reference**

|  |  |
| --- | --- |
| **Population** | Total number/the set of all jobs/reports measured |
| **Standard Deviation (SD)** | A statistic that indicates how tightly the data points are clustered around a mean for a given process, which in turn indicates how much variation exists  |
| **Confidence Interval (CI) *or*** **Confidence Level (CL)** | Refers to the confidence interval which determines the probability that the sample will represent the population. Also referred to as margin of error. |
| **Precision (P)** | Refers to the accuracy level, like decimal points, that we would want when determining the quality of the population  |