Robert W. Allen MBB, PMP

www.robertwalterallen.com

Statistical Analysis of Attribute Data Sample Size and Margin-of-Error

Summary

A statistical analysis was performed on medical coding audit data to ensure acceptable mean accuracy and margin-of-error for both CMSHCC and RxHCC coding processes. The analysis also determined adequate coding audit sample size.

Background

Daily audits are performed on a process by which CMSHCC and RxHCC codes are identified from the electronic health records (EHRs). Two or more EHRs per day are audited for each individual such that a quantity of at least 10 - 16 CMSHCC and 6 - 10 RxHCC codes are audited.

Total codes and correct codes are recorded daily and summarized as % accuracy. An example of daily audit tracking data is provided as follows:

Coder	CMSHCC Total	CMSHCC Correct	CMSHCC Accuracy	RxHCC Total	RxHCC Correct	RxHCC Accuracy
Gemma B	11	11	100.0%	11	11	100.0%
Cindy T	10	10	100.0%	9	9	100.0%
Vicky S	14	13	92.9%	8	7	87.5%

TABLE 1
Sample of Daily Audit Results

"Total" and "Correct" codes are provided from which % accuracy is calculated. The difference between total and correct codes includes missed codes, codes that are not supported, and codes with the wrong level of specificity (therefore both measures of yield and accuracy are included in the % accuracy calculation).

Note that, in this particular example (Table 1), a coder realized a low RxHCC accuracy (87.5%). However, this is a small sample size (1 of 8 missed or incorrect), hence the low calculated accuracy. As with any sampling plan, a small sample may contain a defect and subsequently imply a low yield. A weekly roll-up of audit data is therefore analyzed, with a higher sample size more representative of the population of codes.

An example of weekly audit tracking data is provided as follows:

Coder	CMSHCC Total	CMSHCC Correct	CMSHCC Accuracy	RxHCC Total	RxHCC Correct	RxHCC Accuracy
Gemma B	75	75	100.0%	49	49	100.0%
Cindy T	85	85	100.0%	55	55	100.0%
Vicky S	68	66	97.1%	54	53	98.1%

TABLE 2
Sample of Weekly Audit Results

Audit Process and Measurement System Considerations

Accuracy is the extent to which the measures match the standard value ("truth"). Repeatability means that the same operators, coding the same EHRs, identify the same codes every time. Reproducibility means that different operators, coding the same EHRs, identify the same codes every time.

It is assumed the EHRs and corresponding (coding) results used by the auditors represents "the truth". Repeatability and reproducibility were not analyzed independently in this study, however, this measurement variability would be accounted for in the final accuracy roll-up. A lack of accuracy might warrant further analysis of repeatability and reproducibility.

Approach

The approach presented herein uses sample size analysis of attribute data.

Sample Size Analysis

The following formula is designed for population sampling of a stable process.

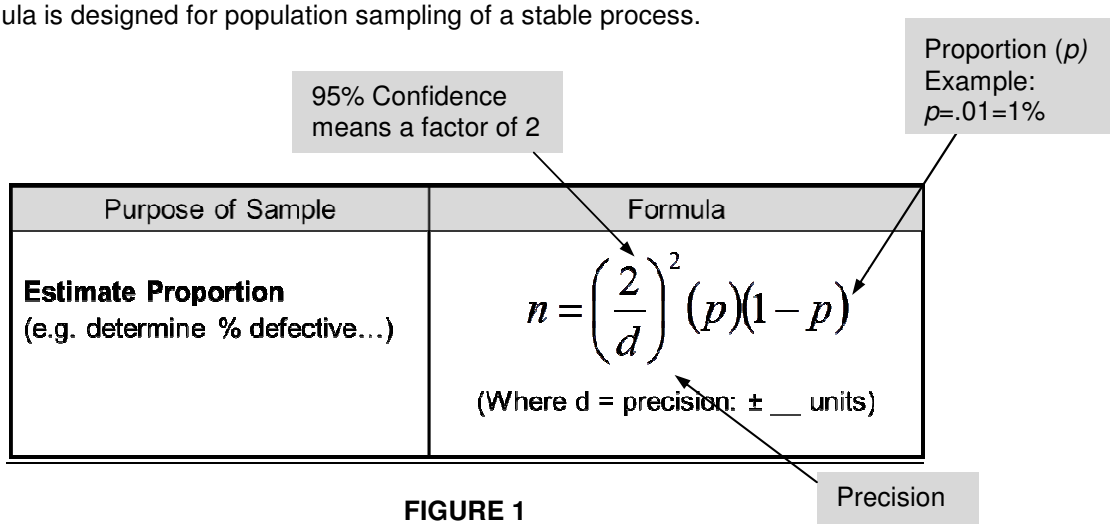

FIGURE 1
Sample Size Formula for Proportion Data

The d value (precision) reflects only the sampling variation, not other sources of variation. We can be confident a range of values contains the average depending on the precision we select. Alternatively, precision (**d**) is how narrow we want the range to be for an estimate of a characteristic and/or is the size of the shift in proportion data we want to detect. Traditional statistics refer to it as "delta" (hence, the d). Precision is equal to half the width of a confidence interval. Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter.

For coding accuracy, assuming we set a 95% Confidence Interval = (98.5%, 99.5% accuracy), this means we are 95% confident the interval from 98.5% to 99.5% contains the true population average % accuracy. The width of the CI = 1%, which means precision = d = 0.5%

Sample size calculations are provided in Table 4, using a selected precision = 0.5%. Also, using the weekly sample size, the corresponding precision was back-calculated.

Sample Size for Estimating a Proportion CMSHCC

- with 95% confidence
- we expect the number of CMSHCC to be 98.7% accurate (p = 0.0126)
- precision selected = 0.5%

$$n = \left(\frac{2}{.005}\right)^2 (.0126)(1-.0126)$$

$$= (160000)(.0126)(.987)$$

$$= 1990$$

Result:
- At the selected precision, a sample size of greater than 1990 should adequately estimate the CMSHCC proportion.
- On any given week, we sample ~2457 CMSHCC codes, which is **more** than the calculated sample size threshold

Estimating a Precision based on CMSHCC Sample Size

- with 95% confidence
- we expect the proportion of defectives to be .0126
- We want the estimate the precision using our weekly sample size ~ 2457 codes/week

$$d = 2\sqrt{\frac{(p)(1-p)}{n}}$$

$$d = 2\sqrt{\frac{(.0126)(1-.0126)}{2457}}$$

$$d = .0045$$

Result:
- Calculated precision = d = 0.45%

Sample Size for Estimating a Proportion RxHCC

- with 95% confidence
- we expect the number of RxHCC to be 98.7% accurate (p = 0.0127)
- precision selected = 0.5%

$$n = \left(\frac{2}{.005}\right)^2 (.0127)(1-.0127)$$

$$= (160000)(.0127)(.9873)$$

$$= 2006$$

Result:
- At the selected precision, a sample size of greater than 2006 should adequately estimate the RxHCC proportion.
- On any given week, we sample ~1723 RxHCC codes, which is less than the calculated sample size threshold

Estimating a Precision based on RxHCC Sample Size

- with 95% confidence
- we expect the proportion of defectives to be .0127
- We want the estimate the precision using our weekly sample size ~ 1723 codes/week

$$d = 2\sqrt{\frac{(p)(1-p)}{n}}$$

$$d = 2\sqrt{\frac{(.0127)(1-.0127)}{1723}}$$

$$d = .0053$$

Result:
- Calculated precision = d = 0.53%

Conclusion

Therefore, based on historical data, we are able to detect a shift in accuracy with a margin-of-error +/- 0.53%. Given our mean accuracy (98.7%) and the Center for Medicare and Medicaid Services (CMS) requirement for coding accuracy (> 95%), it was concluded our margin-of-error was acceptable with our existing sample size.

An advertised accuracy of > 98% was recommended, assuming no shift or drift in the mean.

Assumptions / Considerations

- No shift or drift in the mean is assumed (ongoing monitoring would confirm this).
- Data analyzed is aggregated weekly data. A smaller (daily) sample size, or results from any given individual (coder) may reveal more variability.
- Accuracy of the audit reference standard is assumed to represent “the truth”, and there is no other “truth”.
- Data analyzed is assumed to be random sampled.
- Data analyzed is assumed to be identical to the audit findings.