COGENT
QC SYSTEMS

21 Key Statistical Concepts for
Loan Quality Control

CONTENTS:

- **Introduction**
- **Sample inferences & statistical precision**
- **Random selection**
- **Sample size estimation**
- **Qualitative analysis & defect rates**
- **Random variation & statistical control**
- **Sampling error & non-sampling error**
- **Correlation vs. causation**

Introduction

Robust and appropriate use of statistics in loan quality control (QC) presents significant opportunities for both lowering the costs of QC and for gaining more insight into the quality of loan origination and servicing processes. This paper discusses key concepts in effective *statistical* sampling, which comprises a major part of an organization's sampling protocol. Non-statistical sampling, which is usually an additional requirement for QC but varies widely across organizations, is not addressed here. For more information, be sure to see the companion piece to this paper, "Statistical Sampling in Mortgage Quality Control," available in the Resources section of Cogent's website (www.cogentqc.com).

Sample Inferences & Statistical Precision

The fundamental purpose of sampling for Quality Control is to render judgments regarding the quality of the overall loan portfolio, i.e., to infer general conclusions from the sample's findings. The degree to which those conclusions can be reliably inferred is measured by statistical precision. Keep in mind that the goal of Quality Control is to focus on the forest, not the trees. Accordingly, your objective is not to identify and correct errors or defects in specific loan files, but to use the incidence of such errors to infer conclusions about your loan origination process.

Critical issues:

1. Statistical inference must be based on random selection; the most common error is to draw conclusions from a non-random sample. To avoid this error, you should eliminate all non-random selections from any group used to make statistical inferences to the population.

2. Statistical precision (e.g., of two percent) must be demonstrated on the actual sample defect rate (i.e., the number of loans with defects divided by the number of loans reviewed). If you were

unable to review some of your randomly sampled loan files, then the precision achieved by your process will be degraded.

Random Selection

Random selection means simply that each member of a population, or population segment, has an equal chance of being selected to the sample.

Critical issues:

3. Only true random samples allow a reliable inference to the population. Thus if QC samples and reviews 100 loans, of which 75 are random selections and the remaining 25 represent adverse selections (e.g., EPD reviews, samples drawn from the worst brokers or branches, etc.), then only the 75 random reviews can be inferred to the population.

4. Random selection now usually involves a computer selection using a random number generator. Using a highlighter to select every tenth loan from a list of all closed loans is not a valid random selection technique.

5. Altering your population after selecting a random sample will degrade the quality of your statistical inference. For example, assume that 5,000 June closed loans are added to your QC database and you randomly sample and review 40 June loans in July and August. However, when the July closed loans are added to the database, an additional 300 June loans (perhaps loans that closed on June 30) are also added. This means that when reports of June quality are prepared, the 40 loans sampled might be inferred to the (current) June population of 5,300 – even though none of the 300 subsequently added loans were available for the June random sample.

6. Do not ignore the potential bias of unavailable or missing files. A random selection of, say, 100 loans may result in only 90 reviews, because the remaining 10 files are not available for review. Be extremely leery of making an inference from the 90 reviews completed to the population: your implicit assumption is that the 10 missing files are distributed exactly as the 90 received files. Similarly, the strategy of over-sampling – for example, selecting 120 files for review in the hope that 100 (your desired sample size) will be available – may inject significant bias to any inference drawn from the 100 reviewed loan files.

Sample Size Estimation

A statistical method designed to yield a sample size sufficient to yield a target level of statistical precision (or sampling error). An investor's traditional 10% random sample requirement may be replaced by such a sample, subject to certain requirements. Formula input variables are (1) population size, (2) observed defect rate, and (3) desired precision (e.g., plus or minus 2% at 95% confidence).

Critical issues:

7. Beware of crude tools (homegrown, online or purchased) that use an incorrect or incomplete statistical formula. Common problems observed in the field include:

   - Constant sample sizes, no matter what the population size is.

   - A standard defect rate, often 5%. The correct defect rate to use is your unbiased, unfixed, random defect rate.

8. Do not confuse sample size estimation with actual, measured statistical precision (i.e., the calculated confidence interval). The former is an estimate based on an estimated defect rate; the latter is a calculation that utilizes the actual sample defect rate and is reported after reviews are completed. If your investor requires 2% precision at 95% confidence, you must be able to show that your sample findings support this standard.

Qualitative Analysis & Defect Rates

Qualitative analysis focuses on the incidence of a critical attribute of the loan population. The most critical attribute of a loan is whether - overall - its quality is acceptable or defective. Files with no errors, or minor errors, are deemed acceptable; those with serious errors, or evidence of fraud, are deemed defective. We can then report on sample defect rates, and the inferred maximum and minimum number of defective loans in the population.

Critical issues:

9. Too many QC reports emphasize the frequency of a given error, or all errors, amongst all loans reviewed in a specified period. But since many of these errors are relatively insignificant – they do not affect the overall credit quality of a loan, for example – the report is almost useless to senior management. Such reports may be useful for feedback to and training of production staff, but they do not capture the portfolio in a meaningful way.

10. A lender's definition of "defective" should reflect its specific lending environment, and may not be appropriate to another lender. Within the organization, a consensus should exist as to what defective means.

11. If a reviewed loan is deemed defective, but the critical flaw(s) can be fixed (e.g., a missing promissory note is found after the QC review), do not revise the file assessment to acceptable. The point of sampling is to understand the portfolio – including the majority of loans that will never be reviewed and for which there will be no opportunity to fix errors and omissions.

12. Qualitative analysis (defect rate) allows the maximum efficiency in statistical sample size estimation.

Random Variation & Statistical Control

Random variation in loan QC refers to the observed range of values that a sample will always contain due to sampling error. For example, if 30 loans were reviewed from the production of each of 10 underwriters, we might find that underwriter defect rates ranged from 0% to 10%, with most underwriters clustered around an average defect rate of 3%. Random variation, however, means that chance alone would account for most – and sometimes all – of the variation in defect rates. We can measure the range of defect rates attributable to random variation and conclude that the underwriters falling within that range are essentially equal in terms of loan quality. But if certain underwriters lie beyond this range (the control limit), their defect rate is not due to chance and we can say that the underwriting process reflected by the 10 underwriters is out of statistical control.

Critical issues:

13. A common error made in comparing production units (branches, loan officers, underwriters, etc.) is to calculate the average defect or error rate of the unit under study, and conclude that those with worse than average quality are a problem. But random variation implies that a range of values, both above and below the average value, is statistically insignificant. Thus targeting a broker with a sub-average defect rate for special treatment will not improve overall wholesale quality, and may actual degrade overall quality if the target broker's performance falls within the statistical control limit.

14. Statistical control limits are not arbitrary boundaries or management's target quality range. They are, rather, calculated values based on the standard deviation of the specific sample defect rates under scrutiny.

15. Any attempt to improve the quality of the underwriting process must first correct the quality of the outlying underwriters and bring the underwriting process to a state of statistical control. In other words, if the distribution of underwriter defect rates is not within the control limits, any attempt to improve overall underwriter quality – through training, re-assignment, etc. – is doomed.

Sampling Error & Non-Sampling Error

The total error (i.e., between inferred findings and actual findings) is composed of *Sampling Error* and *Non-Sampling Error*. Sampling error refers to errors introduced by virtue of the fact that samples vary from one to the next. The degree of variation introduces bias into sample inferences. Sampling error can be reduced by increasing sample size (and precision).

Non-sampling errors, on the other hand, are errors of measurement. For example, any process that relies on human judgment will have non-sampling errors introduced as a consequence of variation from one reviewer to another.

Critical issues:

16. The important point here is that there is nothing to be gained by reducing sampling errors below a certain point as compared with non-sampling errors. Total error cannot be significantly reduced unless both sources of error can be controlled. All the statistical precision in the world will not

erase problems related to inconsistent reviewers.

17. If your Quality Control Process ignores non-sampling errors, then significant sources of bias may be introduced into the results. This is the most common shortfall in quality control.

18. One standard approach to determine the degree of non-sampling error is to give the same loan file to multiple Reviewers; if the findings (i.e., Defective or Non-Defective) vary from one Reviewer to another, then there is bias introduced by non-sampling (i.e., measurement) error.

19. Training of Reviewers to produce consistent findings is the most important method for reducing non-sampling errors in Loan Quality Control.

Correlation vs. Causation

Interpretation of Quality Control findings must recognize the difference between correlation and causation. Inferential statistics can identify correlation, but much more sophisticated analysis is required to determine causation. Such sophisticated analyses are beyond the scope of your measurement techniques.

A classic statistical example is the near-perfect correlation between consumption of ice cream and the incidence of drowning. Although one can conclude that ice cream consumption is correlated with drowning, no one would argue that eating ice cream causes one to drown. What is usually happening is that the variables being measured are covariant, i.e., they are both related to a third variable, which is not being measured (in this case, hot weather.)

The best way to minimize problems arising from confusion between correlation and causation is to focus your interpretation on *process*. For example, if your findings indicate that a certain Loan Product has higher defect rates than other product types, then focus on the process by which those loans are originated as compared to the process for other products. If the process is the same, then you can conclude that the variation in defect rates is due to non-sampling bias. Then you can focus on likely sources of non-sampling error (e.g., a Reviewer whose approach is inconsistent with other Reviewers.)

Critical issues

20. The most common error is to conclude that correlation demonstrates a causal relationship. In fact, it almost never does. This is because we are able to measure only a small number of variables that impact the origination process, and much variation may arise from other sources.

21. The best approach to statistical interpretation is to maintain a process focus; your objective should be to identify variations in process that explain variation in defects. To the extent that such process variations do not exist, then focus on non-sampling bias (e.g., inconsistent review criteria.)