

Can your sampling plan see it?

A preregistered test of the CodeMani sampling plan check, in simulation and on hardware

24 September 2026 · codemani.com/sampling.html

Summary

A sampling plan that repeats in step with the process it measures can look at the same spot every time. The data it produces looks normal, and the pattern that matters never shows up. The CodeMani sampling plan check reads a plan before any data is collected and either admits it, or refuses it and names the patterns it cannot see.

We tested that claim two ways, with the verdicts and the predicted catch rates frozen before any test data existed.

- **Simulation, 10 plans in three settings** (an 8-cavity mold, audit sampling of records in date order, and vibration sampled against shaft rotation). For all 83 plan-and-pattern combinations, the catch rate the check predicted from the plan alone matched the measured rate over 2,000 trials each to within 0.6 percentage points. Defects placed in patterns the check named as never seen changed none of the sampled values, so no analysis of that data could have caught them. All three registered predictions passed.
- **Hardware, one robot joint.** Asked to learn a change to a motion it had never seen, the two admitted plans passed and the refused plan failed, in two replicates run in opposite order.

1. The problem

Systematic samples are cheap and easy to run, and they fail quietly when their rhythm matches the process.

- **Audit sampling.** Auditing guidance warns that a fixed sampling interval can coincide with a pattern in the population and produce an unrepresentative sample (AICPA, *Audit Guide: Audit Sampling*; Arens, Elder and Beasley, *Auditing and Assurance Services*).
- **Multi-cavity molding.** Cavity-to-cavity variation is a recognised source of variation, and an interval that lines up with the cavity count can keep sampling the same cavity (Wheeler, "Rational Subgrouping", *Quality Digest*, 2015). Whether it does depends on how the pull aligns with the molding cycle, which is exactly what a plan check can tell you.
- **Fixed-time observation.** A sun-synchronous satellite crosses a given latitude at the same local solar time on every pass, which is why NASA flies MODIS on two satellites with morning and afternoon crossings (NASA, "Aqua and Terra Orbits").

In each case the fix is easy once the blind spot is known. The hard part is knowing, before the data is collected.

2. What the check does

You send the plan, not the data: how you sample (interval, rate, times, sensor positions or a selection rule, and how many samples), the patterns you need to catch, and how small a change must show up, with a noise level and timing uncertainty if you have them.

You get: a verdict (admitted or refused); for a refusal, the named blind spots; a suggested adjustment where one exists; a signed receipt that binds the plan to the verdict; and, as tested here, the catch rate the plan should achieve for each declared pattern.

This paper does not describe how the check works internally. Section 5 describes how you can test it against your own data.

3. Test one: a preregistered simulation

Protocol. For each plan, the check's verdict and a predicted catch rate for every declared pattern were written to a file and committed to version control before any simulation code was run (file SHA-256 22c178c4ea714b6e...). The simulation then injected a defect of six times the declared precision into one pattern at a time, 2,000 times per pattern, added measurement noise, fitted the data with a standard statistical fit that knows nothing about the verdicts, and counted a catch when the injected pattern was both flagged and the most significant pattern found. Three predictions were registered:

1. For every plan and every pattern, the measured catch rate is within 3 percentage points of the predicted rate.
2. Every pattern of every admitted plan is caught at least 95% of the time; every pattern a refusal names as never seen is caught at most 1% of the time.
3. The false-alarm rate with no defect present is within 1 percentage point of the prediction.

Result: P1, P2 and P3 all passed. Largest catch-rate gap 0.6 percentage points; largest false-alarm gap 0.1 percentage points (results file SHA-256 b1206ecab437c37d...). The second half of P2 holds by construction: a defect in a never-seen pattern leaves every sampled value unchanged. One change was made after the run, to the descriptive amplitude readout in 3c (it had read a numerically-zero component as noise); it is not a graded quantity, and every graded number was unchanged.

3a. 8-cavity mold, one dimension measured per inspected part

Parts leave the press in cavity order 1-8, repeating. Measurement sd 0.02 mm. Declared patterns: any single cavity running off-nominal. Budget: each cavity's level known to 0.01 mm (sd).

Plan	Verdict	Blind spots named	Weakest pattern	Predicted catch rate	Measured catch rate	Suggested adjustment
every 8th part, 40 parts	Refused	never sees cavity 2, cavity 3, cavity 4, cavity 5, cavity 6, cavity 7, cavity 8	cavity 2	0.0%	0.0%	every 7th part, 40 parts
every 12th part, 40 parts	Refused	never sees cavity 2, cavity 3, cavity 4, cavity 6, cavity 7, cavity 8	cavity 2	0.0%	0.0%	every 11th part, 40 parts
every 7th part, 24 parts	Refused	too few samples for the precision asked	cavity 7	95.5%	94.9%	every 7th part, 40 parts
every 7th part, 40 parts	Admitted	none	cavity 5	99.9%	99.9%	

Pulling every 8th or every 12th part looks like a reasonable plan. The first inspects one cavity and the second two, so a bad cavity elsewhere is never seen, however many parts are pulled. Every 7th part reaches all eight cavities; with 24 parts it falls just short of the precision asked for, and with 40 it meets it.

3b. Systematic audit sample from records in date order

20 records per business day in date order, so the weekday repeats every 100 records. Per-record error amount sd 1.0 (normalised). Declared patterns: an error level specific to one weekday. Budget: each weekday's level known to 0.35 (sd).

Plan	Verdict	Blind spots named	Weakest pattern	Predicted catch rate	Measured catch rate	Suggested adjustment
every 100th record, 60 records	Refused	never sees Tuesday, Wednesday, Thursday, Friday	Tuesday	0.0%	0.0%	every 98th record, 60 records
every 50th record, 60 records	Refused	never sees Tuesday, Thursday, Friday	Tuesday	0.0%	0.0%	every 49th record, 60 records
every 37th record, 60 records	Admitted	none	Thursday	100.0%	99.9%	

An interval equal to one week of records samples only Mondays; half a week samples Mondays and Wednesdays. Errors concentrated on any other weekday cannot appear in the sample at all.

3c. Vibration sampled against shaft rotation, engine orders 1-12

Samples at a fixed rate relative to shaft speed. Measurement sd 1.0 (normalised). Declared patterns: vibration at any engine order 1 to 12 (amplitude and phase). Budget: each order's components known to 0.5 (sd).

Plan	Verdict	Blind spots named	Weakest pattern	Predicted catch rate	Measured catch rate	Suggested adjustment
24.00 samples per revolution, 10 revolutions	Refused	sees only part of order 12	order 12	95.2%	95.0%	24.02 samples per revolution, 10 revolutions
24.02 samples per revolution, 2 revolutions	Refused	too imprecise on order 12	order 12	88.5%	88.0%	24.17 samples per revolution, 2 revolutions
24.50 samples per revolution, 4 revolutions	Admitted	none	order 1	100.0%	100.0%	

This is the subtle case. At exactly 24 samples per revolution, order 12 is only half visible, so a vibration there is still flagged most of the time, but its size is misread: the median reading was 72% of the true amplitude, and 33% of readings were below half the true size, which is how an alarm limit gets missed. At 24.02 samples per revolution over two revolutions every order is technically visible, but order 12 is so imprecise that its median reading was 141% of the true size, dominated by noise. The check refuses both plans for those reasons and names order 12 in each.

4. Test two: a robot joint on hardware

The second test used a servo joint on a small humanoid robot running a repeated arm motion. A known change was added to the motion, constructed to vanish at each of the refused plan's nominal sampling positions (the actual reads landed slightly off those positions, where the change was small but not zero). Each plan then collected fresh samples and had to predict an independent set of probe readings, taken before any plan ran. The pass threshold (a prediction error score of at most 4.0), the plans, and the expected verdicts were frozen before the run. The run was repeated with the plan order reversed.

Plan	Check verdict	Replicate A score (pass $\leq$ 4.0)	Replicate B score (pass $\leq$ 4.0)
No new samples (previous calibration only)	not applicable	26.83 (FAIL)	25.56 (FAIL)
Plan A1	Admitted	3.67 (PASS)	2.68 (PASS)
Plan A2	Admitted	3.90 (PASS)	1.86 (PASS)
Plan R	Refused	38,958 (FAIL)	55.32 (FAIL)

All registered predictions held in both replicates: the previous calibration alone failed, both admitted plans passed, and the refused plan failed. In a follow-up analysis chosen after the runs (and labelled as such), the refused plan's data could recover only 8 of the 25 declared independent patterns, which is the number the check had predicted. Measured as a share of the previous calibration's excess error removed (replicate A / B), the admitted plans removed 78% / 86% and 76% / 93%; the refused plan removed 26% / 28%. How large a refused plan's failure score gets depends on the fitting method used; only the fail was registered.

Scope: one motion, one joint, one kind of change, two replicates in one evening. The margins for the admitted plans were thin in replicate A and comfortable in replicate B.

5. Limits, and how to test us

- The check answers only for the patterns, noise and timing you declare. A pattern outside that list is outside the answer.
- It assumes the plan is carried out as written. Timing drift or jitter beyond what you declare changes the answer.
- The simulations use known models; real processes are messier. The hardware test is one joint on one robot.
- It is a check of a plan. It is not an engineering sign-off, an audit opinion or a compliance certification.

How to test us. Send a plan you have already run and the patterns you care about. We return the verdict, the named blind spots and the predicted catch rates, signed, before you look at your data again. Then compare. The first plan is checked at no charge while we are in early access.

plans@reply.codemani.com · codemani.com/sampling.html

© 2026 Sfiniti AI. CodeMani is a product of Sfiniti AI, Inc. Simulation files: verdicts SHA-256 22c178c4ea714b6e216b625f93c4e7628388cc659c2dcc65e373581d833ab632; results SHA-256 b1206ecab437c37d5a5d9a8efda240892c2a635bb3adf4420b4d8bb80d2221cc.