

## TEACHER AND STUDENT RESOURCE

# Simulation, Confidence and Evidence



## NC Math 4 • NC.M4.SP.2

### Statistics and Probability

Seven instructional sections with 30 multiple-choice and 30 open or performance tasks, teaching scripts, original mathematical visuals, generous workspaces, and worked keys.

# Free sample | Contents

NC.M4.SP.2 | NC Math 4 mathematics

This sample contains 22 complete selected pages from the 100-page resource, plus this sample contents page. Original printed page numbers are retained. Gaps are intentional.

Included original pages: 1-18, 33-34, 63, 81

## Representative activities

M01: sample PDF page(s) 19; original page(s) 18

M02: sample PDF page(s) 19; original page(s) 18

O01: sample PDF page(s) 20; original page(s) 33

O02: sample PDF page(s) 21; original page(s) 34

Selected teacher prompts and keys: original page(s) 63, 81

## Using this sample

Read the preparation and access guidance before using a selected activity.  
Keep teacher keys and private preparation pages separate from student copies.  
Use the original resource contents to see the scope of the complete packet.

## Complete paid download

The complete resource includes the full PDF task bank and teaching materials.  
Its required companion lab is included with the complete paid download.  
The sample does not include that complete companion archive.

## ALIGNMENT

## Standard and learning targets

Apply informal and formal statistical inference to make sense of, and make decisions in, meaningful real world contexts.

**T1 NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

**T2 NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

**T3 NC.M4.SP.2.3 State hypotheses, check conditions and implement a one-proportion z-test with a contextual conclusion.**

NC.M4.SP.2.1 Design a simulation to create a sampling distribution that can be used in making informal statistical inferences.

NC.M4.SP.2.2 Construct confidence intervals of population proportions in the context of the data.

NC.M4.SP.2.3 Implement a one proportion z-test to determine if an observed proportion is significantly different from a hypothesized proportion.

Each item names its own target. The bank supports multiple sessions; do not assign all 60 tasks in one sitting.

## NAVIGATION

# Choose evidence deliberately

**Warm-up: M01-M04 and O01-O04**

**I Do: M05-M08 and O05-O08**

**We Do: M09-M13 and O09-O13**

**Collaborative practice: M14-M18 and O14-O18**

**Independent practice: M19-M23 and O19-O23**

**Small-group reteach: M24-M27 and O24-O27**

**Exit ticket: M28-M30 and O28-O30**

T1: M01, M04, M07, M10, M13, M16, M19, M22, M25, O01, O04, O07, O10, O13, O16, O19, O22, O25, M28, O28

T2: M02, M05, M08, M11, M14, M17, M20, M23, M26, O02, O05, O08, O11, O14, O17, O20, O23, O26, M29, O29

T3: M03, M06, M09, M12, M15, M18, M21, M24, M27, O03, O06, O09, O12, O15, O18, O21, O24, O27, M30, O30

## DAILY LEARNING TARGETS

# A multi-session teaching path

## Session 1 Diagnose and model

Learning target: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.

Learning target: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.

Use warm-up evidence to choose explicit I Do models. Collect predictions before revealing the worked solution.

## Session 2 Compare and explain

Learning target: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.

Learning target: NC.M4.SP.2.3 State hypotheses, check conditions and implement a one-proportion z-test with a contextual conclusion.

Use selected We Do and collaborative tasks. Ask each learner for an initial representation, then compare methods and checks.

## Session 3 Transfer independently

Learning target: Apply the objectives with explicit restrictions, checks and contextual interpretations.

Choose independent items across every objective. Use small-group tasks for a documented gap and a fresh exit task to test transfer.

## PLANNING

# Materials, access and pacing

## Materials

Use a calculator with normal CDF and inverse-normal functions, plus a spreadsheet or the supplied Python 3 standard-library simulation lab. The companion archive contains `inference_lab.py`, eight replicate CSV files and `simulation_results.json`. Keep the model probability, sample size, replicate count and random seed in every simulation record. Graphs of sampling distributions must label the sample statistic, not an individual outcome.

## Prerequisites

Proportions, complements, random sampling, standard deviation, square roots, normal-curve areas and careful reading of conditional statements.

## Pacing

Use a short diagnostic warm-up, a model and shared task in one session; collaborative and independent reasoning in later sessions; then targeted reteaching and an exit task. Select two to four tasks at a time based on evidence. Each section's full bank is reusable, not mandatory in one lesson.

## Neutral access

Read directions aloud without interpreting the graph, enlarge a page, supply a writing or pointing response method, and allow a learner to dictate an explanation. Keep axes, units, data and mathematical demand unchanged.

## Separate instruction from evidence

Keep the learner's first response. Record answer-supplying cues separately from access supports. A completed teacher graph or worked key is for discussion after independent work, never a student construction template.

## Review convention

All contexts and simulation populations in this packet are invented. Write  $p$  for a population proportion,  $\hat{p} = x/n$  for a sample proportion, and  $p_0$  for a hypothesized value. Use the introductory Wald interval  $\hat{p} \pm z^* \sqrt{[\hat{p}(1-\hat{p})/n]}$  only when justified. Use the null standard error  $\sqrt{[p_0(1-p_0)/n]}$  in a one-proportion z-test. Our conservative classroom approximation check requires at least 10 expected successes and failures for a test, and at least 10 observed successes and failures for this interval. Random selection and independence must also be justified; for sampling without replacement, use  $n \leq 10\%$  of the population. These are instructional checks, not a guarantee of exact coverage. Do not silently use the Wald method for sparse or boundary data. Report probabilities to about four decimals and proportions to about three decimals unless asked otherwise.

## TECHNOLOGY LABORATORY

# Simulation laboratory: what repeats?

## A repeatable protocol

One replicate is one whole random sample followed by one calculated statistic. For a binary model with probability  $p$ , generate  $n$  independent uniform values in  $[0, 1)$ ; code a success when a value is less than  $p$ . Record the success fraction once per sample. Repeat the entire sample process  $B$  times. A fixed seed makes a particular run reproducible; it does not make the simulation exact.

## Supplied runs

A:  $p = 0.40$ ,  $n = 25$ ,  $B = 5000$ , seed 201. B:  $p = 0.40$ ,  $n = 100$ ,  $B = 5000$ , seed 202. C: null  $p = 0.50$ ,  $n = 40$ ,  $B = 10000$ , seed 203. D: empirical resampling from 62 ones and 38 zeros,  $n = 100$ ,  $B = 5000$ , seed 204. E and F: sample means from a population with value 0 with probability 0.90 and value 10 with probability 0.10;  $n = 5$  and 50,  $B = 5000$ , seeds 205 and 206.

## Coverage and bias runs

G:  $p = 0.40$ ,  $n = 100$ ,  $B = 5000$ , seed 207; construct a 95% Wald interval for each replicate and count intervals covering 0.40. H: a biased selection frame yields a success probability of 0.70 even though the intended target population proportion is 0.50;  $n = 100$ ,  $B = 5000$ , seed 208. This deliberately separates sampling variability from selection bias.

## Technology choices

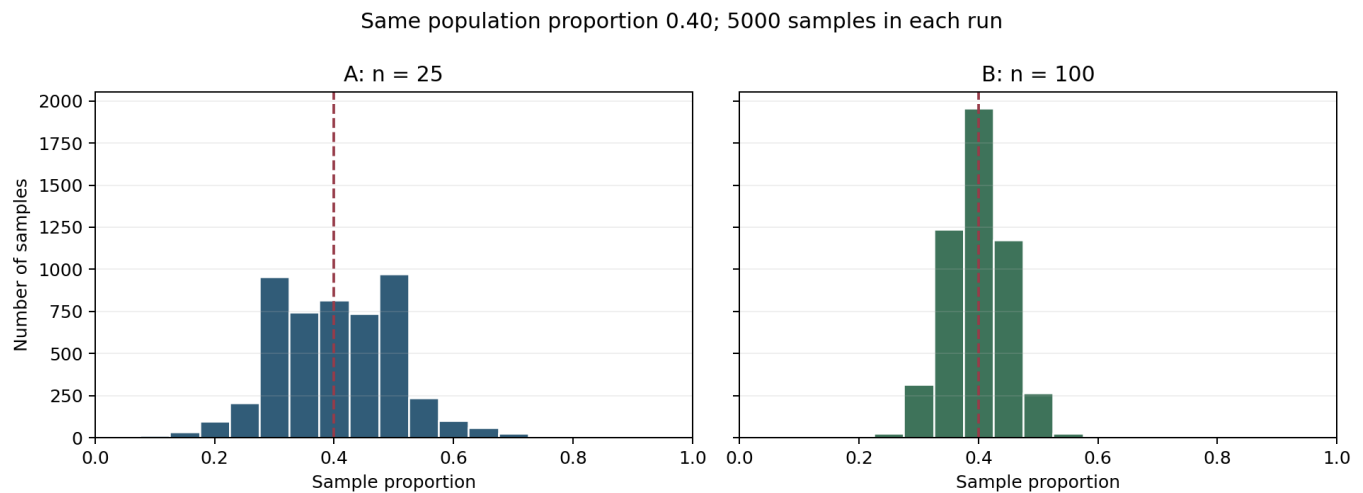
Run `inference_lab.py` with Python 3 to regenerate the CSV files and summary JSON. Alternatively, use spreadsheet random numbers and freeze/save your run before analyzing it. New seeds or tools can produce different counts. For a simulated tail probability, specify the tail rule before counting. Compare observed and simulated statistics on the same scale.

TECHNOLOGY LABORATORY

# Two sampling distributions

## Reference display

These are supplied simulation results, not graphs of individual 0/1 responses. Each histogram uses 5000 sample proportions. Both models have  $p = 0.40$ . The blue display has  $n = 25$  and the green display has  $n = 100$ . Common horizontal and vertical scales make variability comparable.



Original teaching chart generated from the supplied seeded runs A and B. Counts are numbers of whole samples falling in each displayed bin. Read the displayed shape and spread; use the CSV for exact summaries.

## TECHNOLOGY LABORATORY

# Choosing an inferential method

## Confidence interval

For an adequately sized random independent sample, estimate  $p$  with  $\hat{p}$  and calculate  $SE = \sqrt{[\hat{p}(1-\hat{p})/n]}$ . A 95% Wald interval uses  $z^* \approx 1.960$ . A 90% interval uses 1.645; a 99% interval uses 2.576. Confidence describes the long-run coverage of a procedure under its assumptions, not the fraction of individual outcomes within an interval.

## Hypothesis test

Choose  $H_0: p = p_0$  and a directional or two-sided alternative before inspecting the result. Check expected counts  $np_0$  and  $n(1-p_0)$ . Compute  $z = (\hat{p}-p_0)/\sqrt{[p_0(1-p_0)/n]}$ . With  $\Phi$  the standard normal CDF, use  $1-\Phi(z)$  for  $p > p_0$ ,  $\Phi(z)$  for  $p < p_0$ , or  $2\Phi(-|z|)$  for  $p \neq p_0$ . Compare the p-value with the chosen significance level  $\alpha$ .

## Limits of the methods

The Wald interval can perform poorly near 0 or 1 or with small counts; a degenerate or out-of-range result is not fixed merely by clipping endpoints. Use an appropriate alternative such as a Wilson or exact interval with further instruction. A normal z-test is an approximation; a discrete null simulation or exact binomial calculation can differ. Nonrandom sampling and dependence are design problems that larger samples do not automatically repair.

## Method references

NIST/SEMATECH e-Handbook, test of a proportion:

<https://www.itl.nist.gov/div898/handbook/prc/section2/prc24.htm>

NIST discussion of interval alternatives and small counts:

<https://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm>

Python NormalDist documentation:

<https://docs.python.org/3/library/statistics.html#statistics.NormalDist>

The class uses the conservative count threshold 10; this is not presented as NIST's stated threshold.

## TEACHER MODEL

## Separate the population, sample and simulated world

**A random sample has 28 successes in 40 trials. To evaluate a fair binary model informally, what should the simulation repeat and what should it count?**

Say: “I will keep the null probability at 0.50 because that is the model being evaluated, not use the observed 0.70. One simulated sample has 40 independent binary trials and contributes one proportion. I repeat the whole sample many times. The observed distance from 0.50 is 0.20, so for a two-sided comparison I count  $\hat{p} \leq 0.30$  or  $\hat{p} \geq 0.70$ , including equality. The supplied run produces 171 such samples out of 10000, or 0.0171. That is unusual under this model, but it is not the probability the null is true and it cannot repair a biased real-world sample.” Ask learners to identify which quantities stay fixed and which change from replicate to replicate.

## REUSABLE REASONING GUIDE

# Represent, justify, interpret

Name: \_\_\_\_\_ Date: \_\_\_\_\_

**Givens, quantities and restrictions****Representation and calculation****Justification or check****Conclusion, domain and units**

## WARM-UP

# Teaching script

**Core: 001 • T1: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

Optional bank: M01–M04; O01–O04. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 5–8 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O01, including its graph when present. Invite a first independent response before discussion. Ask, “What do the symbols represent? What evidence supports your claim?”

## Core task

Design a technology simulation for the sampling distribution of  $\hat{p}$  when  $p = 0.40$  and  $n = 25$ . Specify one replicate, at least 1000 replicates, the recorded variable, a labeled graph and two features you would inspect. Do not force every sample to have 10 successes.

## Teacher solution and discussion

For each replicate, generate 25 independent uniform values and code values below 0.40 as successes. Record their success count divided by 25 once, then repeat for at least 1000 complete samples. Save the seed or frozen random draws. Make a histogram or dotplot with sample proportion on the horizontal axis and replicate count on the vertical axis. Inspect the center near 0.40 and the spread, as well as shape and possible extremes. The theoretical standard deviation is  $\sqrt{(0.40 \cdot 0.60/25)} \approx 0.0980$ , but an empirical estimate varies across runs.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small p-value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M01 to compare all alternatives after evidence is collected.

I DO

## Teaching script

**Core: 005 • T2: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

Optional bank: M05–M08; O05–O08. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 10–12 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O05, including its graph when present. Think aloud through the core task. Say, “I will identify the quantities first, then build a representation and check it against the stated domain.” Hide the worked key until after predictions.

### Core task

In a random independent sample of 200 adult members from a large association, 120 prefer an online newsletter. Construct 90%, 95% and 99% Wald intervals. Explain the tradeoff without changing the point estimate.

### Teacher solution and discussion

$\hat{p} = 120/200 = 0.60$ ; successes 120 and failures 80 meet the count check. Assume the population is at least 2000 and the sampling description is appropriate.  $SE = \sqrt{(0.60 \cdot 0.40/200)} \approx 0.034641$ . Using  $z^*$  values 1.64485, 1.95996 and 2.57583 gives intervals approximately (0.5430, 0.6570), (0.5321, 0.6679) and (0.5108, 0.6892). All are centered at 0.60. Greater confidence uses a larger margin and less precision; it does not produce a better point estimate.

### Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small p-value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M05 to compare all alternatives after evidence is collected.

## WE DO

# Teaching script

**Core: 009 • T3: NC.M4.SP.2.3 State hypotheses, check conditions and implement a one-proportion z-test with a contextual conclusion.**

Optional bank: M09–M13; O09–O13. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 10–15 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O09, including its graph when present. Have learners propose the next step, then ask a partner to verify it using the original representation. Record more than one valid explanation.

## Core task

A random independent sample of 150 packages from a population of at least 1500 contains 45 with a recyclable insert. Test whether the population proportion is below 0.40 at  $\alpha = 0.01$ . Show why the sign and tail of  $z$  agree.

## Teacher solution and discussion

$H_0: p = 0.40$ ;  $H_a: p < 0.40$ . Null expected counts are 60 and 90, and the stated design satisfies the packet's independence check.  $\hat{p} = 0.30$ ;  $SE_0 = \sqrt{(0.40 \cdot 0.60/150)} = 0.04$ . Thus  $z = (0.30 - 0.40)/0.04 = -2.50$ . The lower-tail  $p$ -value is  $\Phi(-2.50) \approx 0.00621$ . Because  $0.00621 < 0.01$ , reject  $H_0$ . There is evidence that the population proportion with an insert is below 0.40. A negative statistic is evidence in the direction of the specified less-than alternative.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small  $p$ -value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M09 to compare all alternatives after evidence is collected.

## COLLABORATIVE PRACTICE

# Teaching script

**Core: O14 • T2: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

Optional bank: M14–M18; O14–O18. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 15–20 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O14, including its graph when present. Assign reader, representer, checker and explainer roles. Each learner first makes a private prediction; compare reasoning and resolve differences with the data.

## Core task

A volunteer web poll reports 840 supporters among 1200 responses. Another student calculates a very narrow Wald interval and says that its large  $n$  removes bias. Diagnose the problem and propose a better data-collection and reporting plan without treating a formula as a cure.

## Teacher solution and discussion

The point estimate is 0.70, but self-selection and undercoverage may make respondents systematically different from the target population. A nominal interval based only on  $n$  quantifies a model of random sampling variation, not that selection bias. Define the target population, construct a suitable frame, randomly select people and document follow-up and nonresponse. Report the web result as a description of respondents unless a defensible adjustment or design supports more. A larger sample from the same biased mechanism may be more precisely wrong.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small  $p$ -value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M14 to compare all alternatives after evidence is collected.

## INDEPENDENT PRACTICE

# Teaching script

**Core: O19 • T1: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

Optional bank: M19–M23; O19–O23. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 15–20 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O19, including its graph when present. Provide the printed task and neutral access supports. Preserve the first independent response; do not point to the requested answer or select a method for the learner.

## Core task

Use the supplied biased-frame run H to compare precision and accuracy for a target population with  $p = 0.50$ . Calculate the mean and SD of the simulated sample proportions, then describe what a larger  $n$  from the same frame would and would not fix.

## Teacher solution and discussion

The supplied mean is 0.6994 and the sample SD is approximately 0.04573. Estimates cluster near the frame's success probability 0.70, about 0.20 above the intended population value 0.50. Increasing  $n$  from the same biased frame would usually reduce random spread around 0.70 but would not remove the systematic frame bias. A repair must address how cases are selected or appropriately model the selection, not simply collect more of the same kind of cases.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small  $p$ -value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M19 to compare all alternatives after evidence is collected.

## SMALL-GROUP RETEACH

# Teaching script

**Core: O24 • T3: NC.M4.SP.2.3 State hypotheses, check conditions and implement a one-proportion z-test with a contextual conclusion.**

Optional bank: M24–M27; O24–O27. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 10–15 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O24, including its graph when present. Start with the learner's prior evidence. Rebuild one representation with explicit terms and operations, then use a fresh item to check transfer without the prior cue.

## Core task

A random independent sample of 60 orders has 36 successes. Test  $H_0: p = 0.50$  against the prespecified alternative  $p > 0.50$  at  $\alpha = 0.05$ . Explain why changing to a two-sided alternative after seeing the result cannot make its p-value smaller.

## Teacher solution and discussion

Null counts are 30 and 30; assume an adequate population or independent process.  $\hat{p} = 0.60$  and  $SE_0 = \sqrt{(0.25/60)} \approx 0.064550$ .  $z \approx 1.54919$ , so the upper-tail p-value is about 0.06067. Fail to reject at 0.05. For this positive  $z$ , a two-sided test would double that tail to about 0.12134, not reduce it. Alternatives should be chosen for the scientific question before examining the data; switching tests opportunistically changes the procedure.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small p-value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M24 to compare all alternatives after evidence is collected.

## EXIT TICKET

# Teaching script

**Core: O28 • T1: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

Optional bank: M28–M30; O28–O30. Select tasks across more than one meeting if needed.

Suggested discussion/work time: 5–8 minutes. Adjust for access and evidence; this is not a speed criterion.

Display student page O28, including its graph when present. Ask for a concise claim supported by the representation. Collect the response before sharing a solution; use it to choose the next instructional step.

## Core task

For a population taking value 0 with probability 0.90 and value 10 with probability 0.10, design a simulation of sample means for  $n = 5$  and  $n = 50$ . Run or import E and F, compare the empirical mean and spread, and explain why plotting individual values would answer a different question.

## Teacher solution and discussion

For each sample, independently draw  $n$  population values, add them and divide by  $n$ , recording one sample mean. Repeat 5000 times for each  $n$  and graph the sample means on common horizontal scales. Supplied E has empirical mean 0.9988 and SD about 1.34416; F has mean 1.00304 and SD about 0.42405. Both centers are near the population mean 1, while the larger sample size has much smaller variability. A graph of individual values would contain only 0 and 10 and describe the population outcomes, not how sample averages vary. Different runs may yield slightly different summaries.

## Evidence and response

Before calculating, ask: What is the population parameter? What random process generated the sample? What statistic is repeated in the simulation? What assumption makes the tail area meaningful? Require conditions, a correctly chosen standard error and a conclusion about evidence, not a claim that a hypothesis is certainly true or false. A small  $p$ -value does not measure the size or practical importance of an effect.

Closure: Ask learners to state one feature they verified and one check they used. Use the worked key for M28 to compare all alternatives after evidence is collected.

## MULTIPLE-CHOICE BANK

**M01 / M02 • Select and justify****M01 | Warm-up | T1**

**A simulation estimates the variability of the proportion of successful deliveries in random samples of 100. What should one dot in its sampling distribution represent?**

- A. One delivery coded 0 or 1
- B. The success proportion from one complete sample of 100 deliveries
- C. The combined proportion after pooling every simulated sample
- D. The sample size 100 repeated as a response

Reasoning: \_\_\_\_\_

**M02 | Warm-up | T2**

**A random survey finds 62 of 100 residents favor a park plan. Which quantity is the population parameter to be estimated?**

- A. Exactly 0.62, known without uncertainty
- B. The number 100
- C. The probability that a fixed resident changes their answer
- D. The proportion of all residents in the defined target population who favor the plan

Reasoning: \_\_\_\_\_

## WARM-UP

# 001 • Show your reasoning

**T1: NC.M4.SP.2.1** Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.

**Design a technology simulation for the sampling distribution of  $\hat{p}$  when  $p = 0.40$  and  $n = 25$ . Specify one replicate, at least 1000 replicates, the recorded variable, a labeled graph and two features you would inspect. Do not force every sample to have 10 successes.**

**Representation / calculation**

**Explanation / check**

## WARM-UP

**002 • Show your reasoning**

**T2: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

**A simple random sample of 100 residents from a population of 5000 finds 62 in favor of a park plan. Construct a 95% Wald interval for support. State the parameter, check conditions and interpret the result in context.**

**Representation / calculation**

**Explanation / check**

## WORKED TEACHER KEYS

**M01 / M02**

**M01 • T1: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

**Answer: B. The success proportion from one complete sample of 100 deliveries**

Each replicate contributes one  $\hat{p}$ , so 5000 samples produce 5000 plotted sample proportions.

**Check each alternative**

- A. That is an individual outcome, not a sample statistic.
- B. Correct. Each replicate contributes one  $\hat{p}$ , so 5000 samples produce 5000 plotted sample proportions.
- C. Pooling produces one aggregate, not the distribution of repeated sample proportions.
- D. The sample size is fixed and does not represent the statistic's variability.

**M02 • T2: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

**Answer: D. The proportion of all residents in the defined target population who favor the plan**

The unknown  $p$  concerns the population;  $\hat{p} = 0.62$  is one sample estimate.

**Check each alternative**

- A. 0.62 is the observed sample proportion, not a known population value.
- B. 100 is the sample size.
- C. The survey measures current support, not individual change over time.
- D. Correct. The unknown  $p$  concerns the population;  $\hat{p} = 0.62$  is one sample estimate.

## WORKED TEACHER KEYS

**001 / 002****001 • T1: NC.M4.SP.2.1 Design and run simulations of sample statistics; use the resulting sampling distribution for an informal inference.**

For each replicate, generate 25 independent uniform values and code values below 0.40 as successes. Record their success count divided by 25 once, then repeat for at least 1000 complete samples. Save the seed or frozen random draws. Make a histogram or dotplot with sample proportion on the horizontal axis and replicate count on the vertical axis. Inspect the center near 0.40 and the spread, as well as shape and possible extremes. The theoretical standard deviation is  $\sqrt{(0.40 \cdot 0.60/25)} \approx 0.0980$ , but an empirical estimate varies across runs.

Full evidence: A correct probability mechanism, one statistic per full sample, repetitions, labeled display plan and meaningful center/spread checks. Partial credit: preserve and score a valid intermediate representation or justified step separately from the final accuracy.

**002 • T2: NC.M4.SP.2.2 Construct and interpret confidence intervals for population proportions, checking the sampling design and approximation.**

Let  $p$  be the proportion of all 5000 residents who favor the plan. The selection is random,  $100 \leq 500$ , and observed counts are 62 successes and 38 failures, both at least 10. Assuming individual responses are collected appropriately,  $\hat{p} = 0.62$  and  $SE = \sqrt{(0.62 \cdot 0.38/100)} \approx 0.04854$ . The margin is  $1.95996(SE) \approx 0.09513$ . The interval is approximately (0.5249, 0.7151). We are 95% confident that support in the defined resident population is about 52.5% to 71.5%, using this approximate procedure and its assumptions.

Full evidence: Population parameter, random/independence/count checks, estimated SE and margin, ordered endpoints and a population-level interpretation. Partial credit: preserve and score a valid intermediate representation or justified step separately from the final accuracy.