

Digital Measurement and Quantifying Uncertainty

What a Motionless Sensor Reveals About Noise, Bias, and How Much Data Is Enough

Scientific Inquiry with AI
Student Guide

Detailed Instructions & Course Material

Why This Lesson

Place your phone on a table and leave it completely still. Its gyroscope sensor will still report a stream of changing numbers, and the center of that stream may not be exactly zero. The phone is not rotating, so none of that variation is due to the phone moving. All of it comes from the measurement system itself: electrical and thermal noise in the sensor, the finite resolution of the analog-to-digital conversion, and a small offset of the system.

This is an ordinary property of digital measurement, not a defect in your phone. Every instrument in every laboratory behaves this way, which is why a reported number without a stated uncertainty is not yet a measurement. Before you can trust an instrument to tell you something about the world, you have to characterize what it reports when the physical world is not changing. A stationary phone provides an excellent system to practice that, because you already know the true answer is zero, so every departure from zero is something the instrument added.

Most of us use and benefit from digital measurements continuously without noticing them: the time on a screen, the images on our displays, the streams of information our devices send and receive. Routine as they seem, each one rests on a sophisticated measurement system. The delivery of healthcare, the autonomous control of vehicles, the monitoring of our environment, and the vast internet of things all depend on digital measurements, and every one of those applications depends on someone having established how far a sensor's output can be trusted. Advances in measurement science made by the next generation of scientists and engineers will determine the future impact of these technologies and of innovations not yet imagined.

This is the first lesson in which you make measurements with your own phone sensors, and your phone is a more capable instrument than it may appear. Its gyroscope is a MEMS device, a microelectromechanical system with mechanical structures etched into silicon and read out electrically, from the same family of sensors used in vehicle stability control, medical monitors, and industrial instrumentation. Learning to characterize the one in your pocket teaches you what to ask about any sensor you will ever use: how fast it samples, how small a change it can detect, over what range it behaves predictably, how repeatable its readings are, and how close it comes to the truth.

You will use AI as a co-investigator to compute statistics and build graphs that will allow you to visualize the sensor output in ways that would be far too tedious to explore using a spreadsheet. That speed changes what you can ask, allowing you to concentrate on the conceptual understanding of what the data means.

Learning Goals

Upon completion of this investigation, you should be able to:

Phenomena:

- Explain why a stationary digital sensor produces readings that vary, and separate random scatter from a steady bias offset.
- Explain how a small sample size can hide or distort the shape of a distribution, and what a larger sample can establish about the true distribution shape.
- Distinguish the spread of individual readings from the uncertainty in an estimated mean, and explain why collecting more data improves the second without changing the first.

Experimental Practices:

- Use the sensors on smartphones to characterize various physical phenomena and transfer the associated data to a computer for analysis.
- Direct an AI co-investigator to calculate and visualize mean, median, minimum, maximum, range, variance, and standard deviation, requiring it to state the formula and the mathematical steps involved in the analysis.
- Verify AI-produced statistics against independent checks you can perform without the AI: arithmetic identities, unit consistency, order-of-magnitude expectations, and agreement with the visible scale of the raw data.
- Create and interpret time plots, histograms, a running-mean plot, a Gaussian comparison, and the percentages of readings within ± 1 , ± 2 , and ± 3 standard deviations of the mean.

Habits of Mind:

- Predict before measuring and before prompting, so that every result can be compared against your expectation. This enhances learning for the novice and also allows the experienced investigator to identify anomalies.
- Treat AI output as a candidate result rather than an answer, and revise the prompt when the explanation, units, graph, or calculation is incomplete.
- Use mathematical language to describe the statistical characteristics of a measurement.

What You Will Need

- **A smartphone with phyphox installed** — free.
- **A solid, stable surface** where the phone can sit undisturbed for at least five minutes.
- **A laptop** with access to **an AI assistant**.

- **Quantifying Uncertainty_AI_Mediated_Prelab_Tutor.docx** — you will upload this file to your AI assistant for the pre-class tutoring session.
- **This student guide** and a notebook, physical or digital, for predictions and running notes.

The Core Idea

- A stationary phone has approximately zero physical angular velocity, but its gyroscope readings fluctuate around a small nonzero center. The fluctuation is random noise; the persistent offset is systematic bias.
- Mean and median describe the center of a set of readings; minimum, maximum, and range describe the observed endpoints and span; variance and standard deviation describe the typical spread of individual readings around the mean.
- A larger data set does not make the sensor quieter. It makes the underlying pattern, the estimated center, and the shape of the distribution easier to judge, and it narrows the uncertainty in the mean.
- Statistics computed from a small sample are themselves uncertain. Two honest analysts with 25 readings each can report noticeably different standard deviations from the same sensor.
- **Driving question.** How much data is needed to characterize a phone gyroscope's noise honestly, and how well does that noise follow a normal distribution?

Two Kinds of Uncertainty

Both are present in your data, and they behave differently. This lesson will solidify your ability to characterize both types of uncertainty.

Type	What it is and how it behaves	Where you meet it
Random Noise	Unpredictable, unbiased scatter around the true value. Individual readings land above and below the center with no pattern in time. Quantified by the standard deviation, which does not shrink as you collect more data. What does shrink is the uncertainty in your estimate of the mean value.	Stage 1 class distribution; Stages 2–5; post-class extension
Systematic Bias	A persistent offset that shifts every reading in the same direction. Averaging more readings does not remove it — it converges to the wrong answer more precisely. Detecting it requires knowing something independent about the true value. (In this experiment the true value is 0.000 rad/s.)	Stage 2 center check; Stage 5 axis comparison; report claim

Part 1 — Pre-Class Preparation

~80 minutes

The pre-class work is done individually but you are encouraged to talk to others while you work through the activities. The primary purpose of the pre-class activities is to build a foundational understanding of statistical analysis—including the mathematical vocabulary you will use in class—and to introduce you to making measurements with your smartphone's sensors using the free phyphox app.

A. Predict Before You Measure (~10 minutes)

Before installing anything or reading further, make the following predictions in your notebook. These are predictions, not answers, and you are not expected to get them right.

- A gyroscope measures angular velocity around all three axes (x, y, and z) and phyphox will report it in radians/second. Sketch a graph (rad/s vs time) of what you expect a 10-second trace of gyroscope readings from a motionless phone to look like for one of the axes.
- Will the readings be centered exactly on zero? If not, what would cause an offset?
- You will be examining both small and large data sets. Do you expect the magnitude of the variation of the measured values to change over time? How will the certainty of determining the average value change with more measurements?

Why this step first. A prediction you wrote down is evidence about your own mental model and enhances long-term learning as you adjust the model during investigation.

B. Socratic AI Tutoring Session (~40 minutes)

Upload the prelab tutor document to your AI assistant and work through the session. Explain your reasoning in your own words. If you conduct the prelab via the knowlegewithai website, it will save a summary of your chat that your instructor can view. Otherwise, upload the JSON file, that is generated at the end of the prelab session, manually to the knowlegewithai website.

The tutor works on the foundational ideas and on the mathematics behind each statistical value: what the formula does to the readings, why variance squares the deviations, and why the sample standard deviation divides by one less than the number of readings. If you want something to refer back to during class, ask for a one-page summary before the session ends.

C. Explore Phyphox and use AI to Create a Scatter Plot (~30 minutes)

- Install phyphox and explore using the application. Using the accelerometer sensor, practice starting, stopping, visualizing, and clearing data. Identify as many different features as you can in 15 minutes.
- Collect ~10 s of gyroscope data. Export the data and open it in a spreadsheet to explore the format. Determine the sampling rate of the gyroscope and record it in your notebook. Ask your AI for help if you are not sure how to do this analysis.

D. Review of the In-Class Investigation (~10 minutes)

Skim the stages in Part 2 as preparation for our in-class investigation. The data collection and analysis will be more meaningful if you come in prepared and understanding the basic steps you will be conducting.

Part 2 — In-Class Investigation

~2 hours

The Investigation, Stage by Stage

The class works through the stages together, with each student using their own phone to collect their own data. You begin by building a distribution out of the class itself (Stage 1), then collect and interrogate a small data set (Stages 2–4), and finally use a large data set that will provide a more precise statistical characterization of the patterns in the data (Stages 5–6).

Stage 1 — Build a Human Distribution (~15 minutes)

What to do. Measure the time between two claps using the stopwatch on your phones and compare your value with your classmates. Working together, attempt to line up in a way that best illustrates the distribution of times.

What to look for. The line you are standing in is a distribution. Find its center, estimate its spread, and notice its extremes. Everyone was measuring the same physical event with reasonable care, so the spread is not carelessness — it provides insight into the measurement process itself.

Discuss. Is there a value you would call the true interval between the claps? What would you have to know to decide whether the class center is offset from it? That question is the difference between random and systematic error, and you will meet it again in Stage 2 with a sensor instead of a room full of people.

Stage 2 — Collect and Inspect a Short Record (~15 minutes)

What to do. Place the phone on a stable surface, start a recording of approximately 10 seconds, and do not touch the table while it runs. Export the file, import it into your AI model, and create a scatter plots of the data for all three axes (each axis on different graph and no lines between points).

Predict, then observe. You committed to predictions in the pre-class work. Compare your prediction to the real data: Are the measurements center exactly around zero? How wide is the visible variation in rad/s? Does it look the same on all three axes? Is there any measurements that might be real motion — someone bumping the table — rather than noise?

Why inspect the raw values first. Every check you run later compares a computed result against your sense of the data's scale. If you never look at the raw numbers, you have no independent basis for judging whether a reported standard deviation of 0.02 rad/s is plausible or absurd.

Stage 3 — Make the Mathematics Visible (~20 minutes)

What to do. Ask the AI to calculate the mean, median, minimum, maximum, range, sample variance, and sample standard deviation for the gyroscope measurements on all three axes. Do not accept a value-only answer. Require the formula, a description of the mathematical steps, the units of each result, and an explanation of what each value reveals and what it cannot reveal.

A useful prompt start. "Calculate the statistical summaries for each axis of the attached gyroscope data. Before giving me any numbers, show the formula for each statistic and explain in words how it is generated from the readings. Report the units of every result."

Checks you should run on whatever comes back: range equals maximum minus minimum; the mean and median both lie between the minimum and the maximum; the standard deviation is nonnegative and carries the same units as the readings; the variance carries squared units; the standard deviation is smaller than the range, typically by a factor of two to four for a sample this size. Using the graphs from Stage 2, do a quick check to see if the numbers calculated for each of these parameters look reasonable.

Worked Example: Five Readings, Checked by Hand

The following shows an example of a small data set that could be calculated by hand and compared to the results on an AI given the same starting set of values. We will not do this in class unless our quick check suggests that there might be something incorrect happening in the AI analysis.

The readings. Five consecutive z-axis values in rad/s: 0.0021, -0.0007, 0.0013, 0.0002, -0.0009

Center. Sum = 0.0020, so the mean $\bar{x} = 0.0020 / 5 = 0.00040$ rad/s.

Sorted, the values are -0.0009, -0.0007, 0.0002, 0.0013, 0.0021, so the median is **0.0002 rad/s**. Mean and median differ because the single large positive reading pulls the mean and leaves the median alone.

Endpoints. Minimum = -0.0009, maximum = 0.0021, so range = $0.0021 - (-0.0009) = 0.0030$ rad/s.

Spread. Deviations from the mean: 0.0017, -0.0011, 0.0009, -0.0002, -0.0013. Their squares sum to 6.64×10^{-6} .

Sample variance $s^2 = 6.64 \times 10^{-6} / (5 - 1) = 1.66 \times 10^{-6}$ (rad/s)²

Sample standard deviation $s = \sqrt{1.66 \times 10^{-6}} = 0.00129$ rad/s

Independent checks

- **Deviations sum to zero.** $0.0017 - 0.0011 + 0.0009 - 0.0002 - 0.0013 = 0$. This is true for any correctly computed mean and catches an arithmetic slip immediately.

- **Range identity.** maximum - minimum = 0.0030, matching the reported range.

- **Center is bracketed.** Both 0.00040 and 0.0002 lie between -0.0009 and 0.0021.

- **Units.** The variance is in (rad/s)², the standard deviation in rad/s. A standard deviation reported in squared units is a red flag.

- **Order of magnitude.** For five readings, the range is usually about two to three times the standard deviation. Here $0.0030 / 0.00129 = 2.3$, which is consistent.

What five readings cannot tell you. With $n = 5$, the standard deviation itself is uncertain by roughly a third of its value. This example shows you the arithmetic, not a characterization of a sensor.

A silent error worth catching. Ask the AI which denominator it used for the standard deviation. Dividing the sum of squared deviations by n gives the population standard deviation; dividing by $n - 1$ gives the sample standard deviation, which is what you want when estimating a sensor's noise from a finite record. For the five readings above the two answers are 0.00115 and 0.00129 rad/s — an 11% difference from a choice that is rarely stated. At $n = 30,000$ the difference is negligible, but the habit of asking is what generalizes. Nothing about the output looks wrong; the only way to catch it is to ask which formula was used and recompute one case yourself.

Stage 4 — Statistical Analysis of Small Samples (~20 minutes)

What to do first. Working from the 10-second gyroscope data set, ask for histograms for ten different 50 data point subsets (taken at different times throughout the experiment). Before requesting the histograms decide:

- Which axis will you investigate? Why?
- How does the bin width and axis limits influence your ability to visualize and compare the distribution from the 10 histograms? Allow the AI to choose the initial parameters but be prepared to adjust.
- Predict the shape of each graph before you look at it.

What can you conclude from the AI generated histograms? Does the jagged histogram provide evidence that the underlying distribution is irregular?

What to do next. Graph the running mean for the same ten sets of 50 data points.

What can we learn from this type of graph? How many points does it take for the average to approach the mean?

Stage 5 — Statistical Analysis of Large-Sample (~20 minutes)

What to do. Collect a gyroscope data set containing 30,000 data points – this will take coordination to prevent any undesired movement of the tables. Export the completed 300-second record. Ask for full-record time plots, a moving average to expose slow drift, the full statistical summary, and comparable histogram distributions at increasing sample sizes — 25, 50, 200, 1,000, and the complete record (~30,000). Keep the axes identical across the set so the comparison is visual as well as numerical.

Judgment call. What became more stable as the sample grew, and what stayed about the same? The distinction matters: a quantity that stabilizes was uncertain because your sample was small, while a quantity that holds steady is telling you something about the sensor. Do the three axes show different offsets, different spreads, or both?

Look for drift. A moving average that wanders slowly across the 300 seconds is not random noise. It may be thermal, and it means the offset you measured in the first 10 seconds is not a fixed property of the device. Note whether you see it, and say so in your report either way.

Stage 6 — Test a Normal Model (~15 minutes)

What to do. Ask the AI to overlay a Gaussian model on your full-record distribution. Then ask the AI to create a histogram where the bin sizes are one standard deviation and to calculate the fraction of your measurements falling within one, two, and three standard deviations of the mean. Require it to show the counting logic and the exact interval boundaries it used, then complete the table below from your own data.

Band	Your data	Normal model
± 1 standard deviation		68.27%
± 2 standard deviations		95.45%
± 3 standard deviations		99.73%

Interpret. How closely do your percentages and your distribution's shape match the model? Ask the AI to provide you with an assessment on the agreement and what it might tell you about the sensor.

Summarize the AI assessment of the distribution in your own words.

Reflect: Where Does It Break Down? (~10 minutes)

Discuss in groups of four, then record your own answers.

- What did the 10-second record suggest that the 300-second record confirmed, changed, or complicated?
- Which statistic is most useful for describing sensor noise, and why does the way it is constructed mathematically make it the right choice?
- Where did an AI explanation, calculation, or graph need revision, and how did you decide something was wrong?
- What limits your conclusion: vibration, bias, drift, bin choice, outliers, differences between devices, or sample size? Which of these could you have controlled, and which are properties of the instrument?
- Your phone is not a scientific instrument, but a scientific instrument shows the same behavior. What would you now insist on knowing about any sensor before trusting a measurement it produced?

Part 3 — Post-Class Work

~1.5–2 hours

A. Generate and Audit Your Report (~90 minutes)

Prepare a concise report using your own data, a selected set of graphs, and your written reasoning. You may ask an AI to organize or revise the report, but you must audit it line by line against your notes: every number, formula, unit, graph label, and scientific claim. Assessment emphasizes the evidence and the quality of your audit, not length or polish. The report must include:

- How the 10-second and 300-second records were collected, including the surface, the sampling rate, and any disturbance you noticed or suspect.
- The mean and standard deviation with units, interpreted rather than merely reported, with random noise distinguished from systematic bias.
- An explanation in natural language of how the mean, minimum, maximum, range, variance, and standard deviation are generated mathematically from the readings.
- A small set of graphs that directly support your claims. A graph that is not referenced in your argument does not belong in the report.
- The completed $\pm\sigma$ table and a qualified claim about how well the normal model describes your data. Include the histogram for this analysis.
- Your three pre-class predictions alongside what you found, and an account of which specific graph or calculation changed your thinking.
- One useful AI contribution, one output you checked and revised, and one limitation that remains in your conclusion.

C. Retrieval Practice / Course Review (~30 minutes)

Spend approximately 30 minutes on integrated course review using the customized AI-enhanced retrieval tool, which supports spaced repetition and interleaving. Items from this lesson cover random versus systematic error, the construction and meaning of each statistic, why sample size affects the certainty of an estimate but not the noise of an instrument, the $\pm\sigma$ bands of a normal distribution, and the practice of verifying an AI-produced calculation against an independent check.

Everything You Turn In

Your deliverables are listed below with the time they should be completed. The prelab items must be finished before the in-class session, because the in-class work depends on them. The **bold** items will be **turned in**; the others should be available to show your instructor on request.

Due before class, with your prelab

- Your three written predictions about the stationary gyroscope trace.
- **The complete prelab tutor chat log.**

- Confirmation that phyphox records and exports successfully, plus your phone's sampling rate.

Due at the end of the in-class investigation

- Both gyroscope data files: approximately 10 seconds and approximately 300 seconds.
- Your prompts and the follow-up prompts used to correct, clarify, or extend the AI's output.
- The statistical summaries for both records, with units, and the independent checks you ran on them.
- The graph set: time plots, subset histograms, running mean, moving average, the sample-size comparison, and the Gaussian overlay.
- The completed $\pm\sigma$ table and your one-sentence bounded claim about the sensor's noise.

Due after class

- **The audited report, with your corrections marked.**
- The extension: your prediction, the block-mean plot, and your explanation of whether the $1/\sqrt{N}$ relationship held.
- Your answers to the reflection prompts.
- **30 minutes of practice with the retrieval tool.**