Writing Class 3: Assumptions, Validation, and Sensitivity Analysis

Mathematical Modeling Lecture

Mathematical Modeling

Kenneth Sok Kin Cheng

Session Route

Launch	Build	Test
Question Motivating context, key vocabulary, and the first modelling decision.	**Method** Worked example, diagram or table, calculation, and parameter meaning.	**Critique** Checkpoint practice, limitation check, and communication of results.

Columns are used where comparison helps; explanation slides stay clean and spacious.

Why These Three Sections Win Papers

Core explanation, worked reasoning, and application

Why These Three Sections Win Papers

Assumptions, Validation, and Sensitivity Analysis are the three sections that:

- Together count for $15–20\%$ of the score (the *Validation & Sensitivity* pillar).
- Are the most commonly weak in first-time-team papers.
- Are the cheapest to fix — they require almost no additional mathematics, only structured thinking and careful writing.

The combined leverage is large. A team that can write these three sections at **Outstanding-paper quality** can move from Honourable Mention to Meritorious, or from Meritorious to Outstanding, on these alone.

Note: A useful frame: the three trust questions

Judges read these sections to answer three trust questions about your work:

1. **Assumptions:** *Did you think about the real world before you wrote equations?*
2. **Validation:** *Are your results plausible at all, or are they nonsense in disguise?*
3. **Sensitivity:** *Are your results stable enough to act on, or are they an artifact of your parameter choices?*

A confident ”yes” on all three is the signal that distinguishes a publication-quality paper from a homework exercise.

Assumptions: The Three-Part Structure

Core explanation, worked reasoning, and application

Assumptions: The Three-Part Structure

We met this structure briefly in W1. Here we develop it fully.

Method: Method: The Three-Part Assumption

Every assumption in your paper should have three parts:

1. **Statement (What).** What exactly are you assuming? Be specific; an assumption is a claim that can be true or false.
2. **Justification (Why reasonable).** Why is this assumption appropriate? Cite literature, domain knowledge, simplification rationale, or empirical observation.
3. **Implication (What it does for the model).** What does this assumption let you do mathematically? What does it cost you in realism?

Worked Example: Example 1: Three-Part Assumptions for an SIR Epidemic Model

Assumption 1. The total population $N = S + I + R$ remains constant during the modeled period.
Justification: The disease in question (a seasonal influenza) is rarely fatal, and the modeled period is two months, far shorter than demographic timescales. Births, deaths from other causes, and migration thus produce $< 0.5\%$ change in N and can be neglected.
Implication: Reduces the system from three coupled differential equations to two ($S' = -\beta SI/N$, $I' = \beta SI/N - \gamma I$, with $R = N - S - I$). This is the standard SIR reduction.
Assumption 2. The population mixes homogeneously: every infected person is equally likely to contact every susceptible person.
Justification: For a single school community of $\sim 1{,}000$ students in continuous daily contact, this is approximately reasonable, though weaker than for general populations. Network-based models would refine this.
Implication: Allows the bilinear infection term $\beta SI/N$. Without it, we would need a contact matrix or network model, increasing model complexity by an order of magnitude.
Assumption 3. The recovery rate γ is constant (i.e., individuals stay infectious for an exponentially distributed duration with mean $1/\gamma$).

Worked Example: Example 1: Three-Part Assumptions for an SIR Epidemic Model

Justification: Empirical data on influenza recovery times approximate a gamma distribution; the exponential is a one-parameter simplification. The mean is well-measured at approximately 5 days, so $\gamma \approx 0.2$/day.
Implication: Makes I's dynamics first-order, supporting closed-form expressions for the basic reproduction number $R_0 = \beta/\gamma$. A more realistic but more complex alternative would be a gamma-distributed infectious period.

Note: Format conventions for the Assumptions section

- **Number** the assumptions: "Assumption 1.", "Assumption 2.", etc.
- **Bold or italicise** the three parts (Statement, Justification, Implication) for visual scanning.
- Aim for **4–7 assumptions** for a typical paper. Fewer than 4 suggests you have not thought about the modeling; more than 8 suggests you are padding.
- Order assumptions from **strongest to weakest** (most certain first). Judges read down the list with diminishing patience; do not bury your best on page 4.

Worked Example: Example 2: Strong vs Weak Assumptions

Weak: *We assume our model is reasonable.*
Diagnosis: Not a claim. Says nothing.
Weak: *We assume there are 7 days in a week.*
Diagnosis: Trivially true. Wastes space and signals padding.
Weak: *We assume the function is differentiable, integrable, and continuous.*
Diagnosis: Mathematical convenience assumptions stated together. Not a modeling assumption.
Strong: *We assume the daily car arrivals at the toll booth follow a Poisson process. This is supported by the chi-squared goodness-of-fit test ($p = 0.34$) on the 30-day data from the highway authority, and it lets us use exponential inter-arrival times and standard $M/M/1$ queue results.*
Diagnosis: Specific, justified, and connected to model consequence. ✓

Warning: Warning: Distinguish Assumptions from Definitions

- "We define the utilisation as $\rho = \lambda/\mu$" is a *definition*, not an assumption.
- "We assume utilisation is below 1" is a *model constraint*, not really an assumption either (it is needed for stability).
- "We assume that customers' arrival pattern is independent of the service rate" *is* an assumption, because it could be empirically false (longer waits might deter arrivals).

Definitions go in the Variables/Notation section. Assumptions are claims about the world that you could in principle test.

Method: Method: A Checklist of Assumption Categories

Use this list to make sure you have not omitted a relevant assumption:

1. **Boundary / scope:** What system, time period, geography are we considering?
2. **Conservation:** What stays constant? (population, energy, money, mass)
3. **Homogeneity:** Are individuals/units treated as interchangeable?
4. **Independence:** Are events / individuals statistically independent?
5. **Distributional:** What probability distributions are assumed?
6. **Functional form:** Linear, logistic, exponential growth?
7. **Stationarity:** Do parameters change over time?
8. **Equilibrium:** Are we at steady state, or in transient?
9. **Behavioural:** How do humans/agents respond? (Rational? Myopic?)
10. **Data quality:** Is the input data accurate, complete, representative?

Validation: Proving the Model is Not Nonsense

Core explanation, worked reasoning, and application

Validation: Proving the Model is Not Nonsense

Validation answers: *Is your model output plausible?* It is distinct from sensitivity analysis (which asks how output changes with input). Validation comes *before* sensitivity in a well-written paper.

Method: Method: The Five Standard Validation Checks

For each model in your paper, perform as many of the following as feasible:

1. **Limiting-case check.** What does the model predict in extreme cases that you can solve in your head? (Zero customers, zero growth, no friction, etc.)
2. **Conservation check.** If the model should conserve something (population, money), does it?
3. **Dimensional check.** Are the units of every equation consistent?
4. **Literature check.** Does the result agree with published values for similar systems?
5. **Cross-method check.** Does an alternative method (e.g., simulation vs. analytical solution) give the same answer?

Worked Example: Example 3: Limiting-Case Validation for an EOQ Model

The Economic Order Quantity formula $Q^* = \sqrt{2Dk/s}$ predicts the optimal order quantity given annual demand D, ordering cost k, and holding cost s.
Limiting case 1: $k \to 0$ **(no ordering cost).** Then $Q^* \to 0$: order infinitely often in infinitesimally small batches. ✓(Matches intuition: if ordering is free, never hold inventory.)
Limiting case 2: $s \to 0$ **(no holding cost).** Then $Q^* \to \infty$: order all annual demand at once. ✓(Matches intuition: if storage is free, minimise the number of orders.)
Limiting case 3: $D \to 0$ **(no demand).** Then $Q^* \to 0$: do not order. ✓
All three limiting cases check out, increasing confidence in the formula.

Worked Example: Example 4: Cross-Method Validation for a Queuing Model

Our queueing analysis gives the average wait time in an $M/M/1$ queue with $\lambda = 12/\text{hr}$ and $\mu = 15/\text{hr}$ as:

$$W_q = \frac{\rho}{\mu(1-\rho)} \text{ where } \rho = \lambda/\mu = 0.8.$$

$$W_q = \frac{0.8}{15 \cdot 0.2} = 0.2667 \text{ hr} = 16 \text{ min}.$$

We cross-check this against a discrete-event simulation of 10^5 customer arrivals, implemented in Python (`simpy`). The simulation gives average wait $W_q^{\text{sim}} = 15.94 \pm 0.21$ min (95% CI).
The analytical formula and simulation agree to within 0.4%, well inside the simulation's confidence interval. ✓

Worked Example: Example 5: Literature Comparison

Our model predicts that a single bicycle highway between the city centre and the Eastern residential district will shift 18% of commuters from cars to bicycles. We compare against three published European deployments:

Source	City	Mode shift	Notes
Pucher & Buehler (2008)	Copenhagen	22%	After 10 years of network expansion
Hellinga (2017)	Utrecht	16%	After single highway opening
Aldred et al. (2019)	London CS3	14%	After single highway
Our model	—	**18%**	Single highway, mid-sized city

Our prediction falls within the range of comparable single-corridor deployments (14–22%) and lies between the values for single-corridor cases (Utrecht, London) and the long-term, network-scale Copenhagen result, as we would expect.

Note: Format conventions for Validation

- Validation appears as a numbered section (often §6 or §7), *not* inside Model Development.
- Each check is a sub-subsection or a clearly labelled paragraph.
- If a check fails, **do not hide it** — discuss what failed and what it implies. Honest discussion of a failed check is far stronger than silent omission.

Sensitivity Analysis: How Robust is Your Answer?

Core explanation, worked reasoning, and application

Sensitivity Analysis: How Robust is Your Answer?

Sensitivity analysis answers: *How does the output change as inputs change?*
This is the single most commonly weak section in first-time-team papers. Many teams write a few sentences claiming their results are “robust”. Strong teams build sensitivity tables, tornado diagrams, or sensitivity heatmaps that constitute a visual narrative on their own.

Method: Method: Three Levels of Sensitivity Analysis

From simplest to most thorough:

1. **One-at-a-time (OAT).** Hold all parameters at their nominal values; vary one parameter at a time ($\pm 10\%$ or $\pm 20\%$); record effect on output. Simple, fast, but misses interactions.
2. **Tornado diagram.** A horizontal bar chart showing the effect of each parameter when varied within its plausible range. Bars sorted by magnitude. Immediately identifies the most influential parameters.
3. **Monte Carlo / global sensitivity.** Sample all parameters jointly from plausible distributions; compute the distribution of outputs. Captures interactions. Best for high-credibility papers.

For competition papers, OAT plus a tornado diagram is usually sufficient. Reserve Monte Carlo for problems where parameter interactions matter (e.g., queuing under load).

Worked Example: Example 6: OAT Sensitivity Table

Our model predicts a baseline of 312 drones needed. We vary each input parameter $\pm 20\%$ from its nominal value and report the effect on the recommendation.

Parameter	Nominal	-20%	$+20\%$	Range
Viewing distance d (m)	200	287 drones	339 drones	$\pm 9\%$
Image height h (m)	50	256 drones	369 drones	$\pm 18\%$
Drone spacing s (m)	1.0	488 drones	217 drones	$\pm 56\%$
Refresh rate f (Hz)	30	312 drones	312 drones	0%

Interpretation. The recommendation is most sensitive to drone spacing s, which dominates the geometric resolution constraint. Image height has secondary effect. Viewing distance has modest effect. Refresh rate has no effect (the model is quasi-static).

Implication for recommendation. Drone spacing should be measured/specified precisely. Refresh rate can be left to operational judgement.

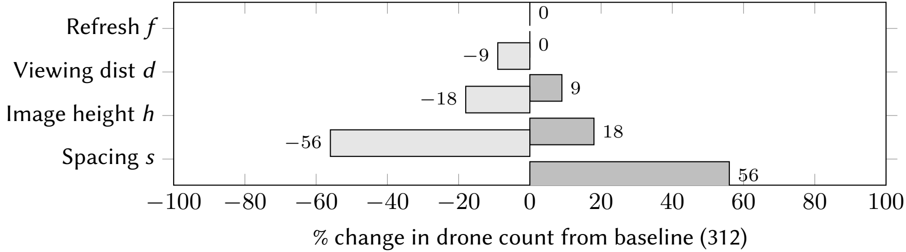
Note: Why tables work better than prose for sensitivity

Compare the table above with: "Our model is relatively sensitive to drone spacing, which causes large changes, and to a lesser extent image height, while viewing distance has only modest effect and refresh rate is essentially neutral."

The table delivers the same information at a glance *and* with numbers *and* with a clearer ranking. Tables almost always beat prose for sensitivity. Use prose to interpret the table, not to replace it.

Worked Example: Example 7: Tornado Diagram

The tornado diagram below visualises Example 6's OAT data:



Sorted by absolute effect; the dominance of s is immediately visible.

Structural Sensitivity: Beyond Numerical Perturbation

Numerical sensitivity tests parameter changes within an assumed model. **Structural sensitivity** tests whether the conclusion survives when the *model itself* changes.

Method: Method: Structural Sensitivity Checks

Ask the following:

1. If we replace the exponential service distribution with deterministic, what happens?
2. If we replace logistic growth with exponential growth, what happens?
3. If we relax linearity in the dose-response curve, what happens?
4. If we add second-order effects to a first-order model, what happens?

Even one structural check, well-executed, signals model maturity.

Worked Example: Example 8: Structural Sensitivity Check

Our queuing model assumes exponential service times (the $M/M/1$ assumption). To test the structural sensitivity of our recommendation, we re-run the analysis with deterministic service times (an $M/D/1$ queue).
Under $M/M/1$: average wait $W_q = 16$ min, recommended staffing $= 3$ baristas.
Under $M/D/1$: average wait $W_q = 8$ min, recommended staffing $= 2$ baristas.
The headline recommendation changes by one barista. We thus report a recommendation **range** of 2–3 baristas with 3 as the upper bound, and we note that more accurate service-time data would allow tighter recommendation.

Putting Them Together: The Trust Chain

Core explanation, worked reasoning, and application

Putting Them Together: The Trust Chain

A great paper builds an **evidence chain**: assumptions → model → validation → sensitivity → recommendation. Each link supports the next.

Method: Method: The Evidence Chain Self-Check

Before submitting your paper, work through this chain backward from your recommendation:

1. **Recommendation:** What am I recommending?
2. **Sensitivity:** Is this recommendation stable to plausible changes in inputs?
3. **Validation:** Has the model that produced this recommendation passed at least one independent check?
4. **Assumptions:** Are the assumptions behind the model explicit and defensible?
5. **Coverage:** Does the recommendation actually answer the question that was asked?

Break in any link of the chain weakens the entire paper. Strong papers have all five visible to the judge.

Worked Example: Example 9: A Complete Evidence Chain

Recommendation. The bicycle highway network should be built (47 km, EUR 28M, predicted 18% modal shift).
Sensitivity. Modal shift varies from 14–22% across plausible weather and demographic assumptions; the recommendation is robust within this range. The most influential parameter is winter rainfall frequency.
Validation. The 18% prediction falls within the range of comparable European deployments (14–22%) per three published studies; our value lies sensibly between single-corridor and network-scale data points.
Assumptions. Key modelling assumptions: logit mode-choice form with parameters from European studies; constant winter weather; no major changes to public transit during the implementation period.
Coverage. The original problem asked for network design and commute time reduction; both are answered.

Every link in the chain is short, specific, numbered, and visible. This is what makes the paper trustworthy.

Workshop Exercises

Core explanation, worked reasoning, and application

Exercises: Exercise W3.1: Rewriting Assumptions

The following assumptions are taken from a (constructed) weak paper. Rewrite each using the three-part structure (Statement / Justification / Implication).

1. *We assume the road is straight.*
2. *We assume the population grows logistically.*
3. *We assume customers are rational.*
4. *We assume there is no friction.*
5. *We assume the data is accurate.*

Exercises: Exercise W3.2: Choose the Validation Method

For each of the following models, suggest two appropriate validation checks (chosen from the five in Method 4.1). Justify your choice.

1. A discrete-event simulation of a hospital emergency department.
2. An optimisation model for warehouse layout.
3. A regression model for sales prediction.
4. A network model for transport corridor routing.
5. A stochastic model for inventory under uncertain demand.

Exercises: Exercise W3.3: Build a Sensitivity Table

Consider the following hypothetical setting. A model for daily revenue at an ice-cream shop is

$$R = (a - bT) \cdot p \cdot N$$

where T is the day's temperature (°C), p is the price per cone (\$), N is the number of potential customers, and a, b are calibrated constants. Nominal values: $a = 5$, $b = 0.05$, $T = 25$°C, $p = 4$, $N = 200$, giving $R = (5 - 0.05 \cdot 25) \cdot 4 \cdot 200 = \$3000$.

Build a one-at-a-time sensitivity table varying each of T, p, N, a, b by $\pm 20\%$, and identify the parameter the recommendation is most sensitive to.

Exercises: Exercise W3.4: Diagnose the Evidence Chain

Read any one Outstanding paper. Identify the team's:

- Top-line recommendation.
- Validation checks performed.
- Sensitivity analysis methodology and main findings.
- Key assumptions.

Now trace the chain backward from recommendation to assumptions. Where are the strongest and weakest links?

Exercises: Exercise W3.5: Write a Full Trio

For your most recent modeling project (course or competition), draft:

1. An Assumptions section with 5–6 three-part assumptions.
2. A Validation section using at least two of the five standard checks.
3. A Sensitivity Analysis section with one OAT table and one structural check.

Have a teammate read all three and ask: *do you trust the recommendation more after reading these than you did before?* If the answer is ”yes, much more”, you have succeeded.

What to Carry Into W4

Core explanation, worked reasoning, and application

Note: Class W3 Take-aways

1. These three sections sit at the heart of the *Validation & Sensitivity* pillar (15–20% of score), but their leverage is greater because they distinguish thoughtful from formulaic work.
2. Every assumption needs Statement, Justification, Implication.
3. Validation has five standard methods; do at least two; include cross-method when possible.
4. Sensitivity analysis must include numbers and a table or figure — not just prose.
5. Structural sensitivity (changing the model) is rare and high-value; do at least one when feasible.
6. The evidence chain from recommendation back to assumptions must hold link-by-link.

What to Carry Into W4

In **W4: English Academic Writing Style** we shift from *structure* to *language*: how to write sentences that sound like scientific papers rather than like translations from another language. We address voice, hedging, connectives, and common Chinglish patterns.

Appendix A: Sensitivity Analysis Quick-Start

Core explanation, worked reasoning, and application

Appendix A: Sensitivity Analysis Quick-Start

Step	Action
1. Identify parameters	List every numerical input to your model.
2. Pick a range	For each, choose a plausible $\pm$ range (often $\pm 10\%$ or $\pm 20\%$).
3. Vary one at a time	Re-run the model with each parameter perturbed.
4. Tabulate effects	Record output change for each parameter.
5. Rank by influence	Sort the table by absolute effect.
6. Plot tornado diagram	Visualise the ranking.
7. Discuss	In one paragraph, identify the most/least influential parameters and what this implies for the recommendation.
8. (Optional) Structural	Replace one model component with a different functional form; re-run.

Appendix B: Validation Checklist

Core explanation, worked reasoning, and application

Appendix B: Validation Checklist

For each model in your paper, check:

- ☐ Does the model behave correctly in at least one limiting case?
- ☐ Are quantities that should be conserved actually conserved?
- ☐ Are the units consistent throughout?
- ☐ Does the result match (or fall in the range of) published values for similar systems?
- ☐ Does an alternative method give a comparable result?
- ☐ If a check failed, is that discussed honestly in the paper?

Appendix C: Strong Assumption Templates

Core explanation, worked reasoning, and application

Appendix C: Strong Assumption Templates

Use these as starting points; modify and extend.

- *We assume X is approximately constant during the modeled period [Y], because [reason]. This allows us to [model consequence], at the cost of [realism trade-off].*
- *We assume X follows a [distribution] with parameters estimated from [data source]. We confirm this fit via [test/visual], obtaining [p-value or descriptive metric]. This supports the use of [standard model relying on this distribution].*
- *We assume X and Y are independent, which is justified by [data, theory, or convention in the field]. This allows the joint probability factorisation* $P(X, Y) = P(X)P(Y)$*, simplifying the calculation of [downstream quantity].*
- *We restrict our analysis to the regime [condition]. Outside this regime, model assumption [A] would fail, requiring a [different approach]. We address that regime briefly in Section [Z].*