SAMPLE

Building Tomorrow's Data Scientists

The Estimate's Keeper in the Machine Era

Aiman JoeHanz

Free sample: Introduction & Chapter 1

putrabooks.github.io

Buy the full book:

Carousell carousell.sg/p/building-tomorrow-s-data-scientists-ebook-pdf-1471385062

Ko-fi

Payhip

Contents

Introduction: The Number With a Range Around It

Chapter 1 — The Estimate

Chapter 2 — The Territories

Chapter 3 — The Path

Chapter 4 — The Working Day

Chapter 5 — The Estimate

Chapter 6 — The Doubt

Chapter 7 — The Conversation

Chapter 8 — The Instruments

Chapter 9 — The Year Whole

Chapter 10 — The Long Path

Introduction: The Number With a Range Around It

There are two kinds of people who produce numbers in an organisation, and the difference between them is not skill.

The first produces a figure. It is on a slide by Tuesday, it has a decimal point, and it is discussed as though it were a measurement of something that exists. The second produces an estimate: a figure with a range, a statement of what it was computed from, an account of what was assumed, and a sentence describing the circumstances in which it would turn out to be wrong.

Both people can be technically excellent. The difference is that one of them knows what they have made and the other does not, and the knowledge is the whole of the trade this book describes. Data science, in the sense that survives contact with an organisation, is the discipline of producing estimates and standing behind them. Everything else, the models, the code, the charts, the pipelines of experiments, is machinery in service of that.

An estimate is a claim about the world that has been constructed carefully enough to be examined. That is the thing worth keeping, and it is what this book calls the estimate's keeper: the person in the room who can say how confident we should be, what would change the answer, and what we would have to believe for this to be true.

The pressure against this discipline is enormous and not malicious. Decisions are made in meetings, meetings need numbers, and a range is harder to put on a slide than a point. A practitioner who delivers a confident figure is praised; one who delivers a range with conditions is asked to simplify. Do that four times and the conditions are gone, and the organisation is making decisions on figures that nobody can defend and nobody remembers constructing.

This book argues for three commitments that hold up against that pressure.

Ask the question before you choose the method. A model is one reply to one question, and the question itself must first be phrased before the technique is selected, because each technique answers a different question and most of them answer it in a way that cannot be checked afterwards. The practitioner who arrives with a method and looks for a question will produce something, and the something will not be an estimate.

Separate what was measured from what was assumed. Every figure contains

both, and they are not equally trustworthy. A measured quantity has an error you can estimate; an assumption has a justification you can inspect. Mixing them in one number is how an organisation comes to believe a guess more strongly than a measurement.

Publish the conditions under which you would be wrong. An estimate without its failure conditions is a claim, and claims are believed without qualification. Stating what would falsify the estimate is not modesty; it is the mechanism by which the work can be corrected before it does damage.

The chapters ahead map the ground this work covers, the roads people travel into it, an unremarkable day and an unremarkable year, the craft of estimation done honestly, the specific doubt that comes with a discipline built on uncertainty, the conversations the work cannot avoid, the instruments it leans on, and the long road. They are addressed to a practitioner in the third or fifth year who can build a model and has begun to suspect that the model was never the point.

It was not. The point is being able to say how much of that figure we actually know.

Chapter 1 — The Estimate

What an Estimate Carries

An estimate is not an imprecise number. It is a number with an account attached, and the account is the valuable part.

Four things travel with a defensible estimate. What quantity it is: not asking costs, but the average cost of a support contact, established over a defined window, for a defined population. What it was computed from: the records, the sample, the period, the exclusions. What was assumed: that the population is comparable over time, that the measurement instrument did not change, that the missing values are missing for reasons unrelated to the value. And how it might be wrong: the conditions under which the estimate is not merely imprecise but misleading.

Most numbers that circulate in organisations carry none of these. They arrive as arithmetic and are treated as facts, and the trade this book describes is the discipline of attaching the account before the number is used to decide something that cannot easily be undone.

Two Folklores

Two stories about what a data scientist is circulate widely, and both produce practitioners who are useful in a narrow range and dangerous outside it.

The oracle believes the model is the answer. The craft is the algorithm, the data is raw material, and the output is a prediction that either matches the future or does not. This story tempts because it counts and because occasionally it is true: for a well-posed prediction with plenty of history and a stable world, the machinery works. Its error is that a prediction without a stated confidence and a stated population is not usable in a decision, and a practitioner who has never had to state those has not learned the trade.

The chart whisperer believes the work is persuasion. The analysis exists to move a meeting; the visualisation is the deliverable; the estimate is a device in an argument. This story is attractive because it is rewarded immediately and because the practitioner who can make a chart that convinces is genuinely valuable. Its error is that persuasion and estimation come apart exactly when the decision is expensive, and at that point the practitioner has no defence and the organisation discovers that it has been persuaded rather than informed.

The account that endures both pressures is drab. The assignment is to turn

out estimates that can be examined, to state their limits, and to be the person who says how much is known. Sometimes that convinces, and the convincing is a by-product rather than the purpose.

Three Commitments

Ask the question before you choose the method. The order matters because the method constrains the answer. A decision tree, a regression, an uplift model and an experiment all answer different questions, and choosing the method first means choosing the question by accident. The practitioner who writes the question down in a sentence before opening an editor has already avoided the most common failure in the territory: a technically sound analysis of something nobody asked.

Separate what was measured from what was assumed. A figure assembled from a sample and a figure assembled from a judgment look identical on a slide, and they are not the same kind of thing. The discipline is to keep them apart in the working and to say which is which in the reporting. An estimate whose assumptions are visible can be attacked in the right place, which is how it gets improved rather than discarded.

Publish the conditions under which you would be wrong. This is the hardest commitment and the most protective. It converts an estimate from an assertion into a testable claim, and a testable claim can be checked by somebody who distrusts you. Stating it also tends to improve the estimate, because writing down what would falsify it forces the practitioner to look at the part of the reasoning they were least sure of.

Where the Difficulty Lives

None of the three poses a technical challenge, and the trade is hard for reasons that are social.

The first is that the demand arrives before the design. Somebody needs an answer on Thursday, and a defensible estimate takes three weeks, and the compromise is an indefensible estimate delivered on Thursday and never revisited. The compromise is always locally reasonable and cumulatively fatal, because the indefensible number now exists and will be quoted.

The second is that organisations reward confidence. A practitioner who says this figure is 12,400 with a range of 9,000 to 17,000 and several important caveats is less useful in a meeting than one who says twelve thousand. The reward is immediate and the cost is delayed, sometimes by years, and it is

paid by whoever is in the role when the estimate turns out to have been wrong.

The third is that the discipline's failures are quiet. A model that never gets deployed, an estimate that is ignored, an experiment that was stopped early. None of these produce a visible incident, so nothing teaches the organisation that it should have listened, and the practitioner's warnings accumulate as a private record that nobody consults.

The fourth is that nothing in the work announces completion. A model can always be validated further, an assumption always examined, an uncertainty always narrowed. Practitioners who cannot decide when an estimate is good enough produce perfect analyses of irrelevant questions, and the trade's reputation is made or lost on that judgment.

The Figure That Was Already a Decision

A great many numbers arrive at the practitioner already carrying a decision, and recognising this saves months of unnecessary work.

The marker is that the request specifies a conclusion. We need to show that the new pricing is better. We need the figures that support the expansion. It is not usually dishonest: the requester has a view, the view is probably informed, and the analysis is expected to document it. The difficulty is that an analysis performed to support a conclusion is not an estimate, because the freedom to choose the specification has been spent before the work began.

The response that works is not a lecture about bias. It is to accept the request and reframe the deliverable: here is what the data says, here is the case for the decision, and here is the case against it. A practitioner who delivers both sides is useful to somebody trying to make a defensible argument, and is trusted precisely because the second column exists. One who delivers only the requested side produces something that collapses the first time it is challenged, taking the decision with it.

The second response is to price the risk. If the analysis is scoped to confirm, it cannot detect the case where the decision is wrong, and the requester should know what that costs. Stated once, calmly, with the magnitude, this changes more decisions than any amount of methodological objection.

The Estimate That Was Right and Costly

There is a category of correct estimate that is expensive to have produced,

and practitioners should understand it before they meet one.

The estimate shows that a programme does not work. Or that a celebrated intervention's effect is indistinguishable from zero, or that the metric the organisation has been optimising has been measured in a way that inflated it for a year. All of it is true, well established, and unwelcome, and the practitioner who produces it discovers that being right is not a protected position.

What protects the practitioner is not the strength of the evidence. It is the process: the question circulated in advance, the plan written before the data was examined, the sensitivity table, and, most importantly, the allies who reviewed the work while it was being done. An estimate that arrives as a surprise is a claim by an individual; one that arrives after review is a finding by a group, and organisations treat the two completely differently.

The second protection is framing. A finding that a programme does not work is heard as an accusation; the same finding framed as what we now know about where the effect is largest is heard as progress. Both are the same estimate and the second is not spin: it usually turns out that the effect is real in a subpopulation, and looking for that is better work than defending a null.

The third protection is to be right about the thing that matters. A practitioner who spends a year proving a programme ineffective, when the programme is being ended anyway, has won an argument nobody was having. Establishing what the decision actually is, before the analysis, is what keeps the work relevant.

What an Organisation Believes About Numbers

An organisation holds a set of beliefs about its own figures, and those beliefs are rarely written down and almost never examined.

The first belief is that a number in a system is more trustworthy than a number in a document. This is mostly true and it fails precisely where it matters: a system's number is trustworthy about what the system records and silent about what it omits, and the omissions are where the errors live. A document assembled by a person carries its author's judgment, which is inspectable; a system's figure carries nobody's judgment, which makes it feel objective and makes it uninspectable.

The second belief is that precision indicates quality. A figure with two decimal places looks more considered than one rounded to the nearest thousand, and

the appearance is unrelated to the underlying knowledge. Practitioners who give a precise number because it looks professional have communicated a confidence they do not have, and the communication is not undone by a caveat three paragraphs later.

The third belief is that agreement between two sources confirms a number. Two systems agreeing usually means one was derived from the other, or both were derived from the same input, and the agreement is a statement about lineage rather than about truth. The confirmation only counts when the two sources are independent, and independence is much rarer than the appearance of it.

The fourth belief is that a number that has been used for a year has been checked. In practice the first use is the check, and after that the number is quoted from the previous document, and its provenance decays with each quotation until nobody can say where it came from. This is the mechanism by which an error survives a decade, and it is the reason the trade's characteristic task is not computation but archaeology.

The fifth belief is the most damaging. It is that an estimate and a guess are different only in degree. They are different in kind: an estimate has an account, a guess has a feeling, and the account is what allows the organisation to improve. Where that distinction is not held, the practitioner's careful work and a colleague's hunch compete on equal terms in a meeting, and the hunch usually wins because it is delivered faster and with more confidence.

A practitioner who understands these beliefs can work with them rather than against them. The useful move is not to argue that the organisation is naive. It is to supply, at the moment of use, the two facts that the beliefs omit: what this figure leaves out, and what would have to alter before it came out wrong.

The Range That Kills the Meeting

A range is the honest form of an estimate and it is also the form that most reliably ends a decision, which is a tension the practitioner has to manage rather than resolve.

The mechanism is that a range invites a question the meeting cannot answer: which end of it should we plan for. A point estimate silences that question by fiat, and the silence is comfortable and frequently wrong in one direction. Practitioners who offer ranges without guidance therefore create a vacuum

which is filled by whoever is most confident rather than whoever is best informed.

The version that works pairs the range with a recommendation. The estimate is between nine and seventeen thousand, we would plan for twelve, and if we are wrong it will be at the low end because the population assumed in the upper half has not yet been onboarded. That sentence gives the meeting something to act on and preserves the uncertainty, and it takes twenty seconds longer than the dishonest version.

The second technique is to attach the range to a decision rather than to the quantity. The difference between the two ends does not change which option we should choose, or it does, and if it does here is the additional analysis that would resolve it. This converts uncertainty from an obstacle into a description of what is at stake, which is what a decision maker actually needs.

The third technique is the demand curve. Where an organisation insists on a point estimate, the practitioner can supply one and state what it is conditional on, once, and then stop. Repeated caveating in the same meeting converts a well-founded reservation into a personality trait, and the trait is remembered after the reservation is forgotten.

The underlying judgment is about purpose. Some decisions need a range because the range is the decision, and others need a point because the decision cannot wait for precision. Telling the two apart is the craft, and it is learned by watching what the decision maker does with the answer.

The Meeting Where the Number Was Simplified

Somewhere between the analysis and the decision, a figure is simplified, and the simplification is almost never recorded.

It happens in a preparation meeting, in a slide deck, in a summary paragraph. The estimate becomes a number, the number loses its range, the range loses its conditions, and by the time it reaches the decision the only thing remaining is a quantity that looks like a measurement. Nobody in the chain intended to misrepresent anything; each step removed a qualification that seemed unnecessary for the audience.

The practitioner's instrument against this is the one-page summary from chapter five, circulated before the meeting rather than afterwards. A document that already contains the figure and its conditions travels better than an explanation given verbally and summarised by somebody else. Where

the document exists, the simplification still happens, but it happens in the presence of the original.

The second instrument is to state the single most important condition rather than all of them. Ten caveats are ignored; one is remembered. The practitioner who says, every time, this figure assumes the pilot cohort, and the pilot cohort is not representative, has chosen the condition that matters and repeated it until it is part of how the number is discussed.

The third is to accept simplification where it is harmless. A decision that is robust to the range does not need the range, and insisting on it in that case spends credibility that is needed later. The judgment about which simplifications matter is the trade's practical form of risk assessment, and it is built by watching what decisions do with what numbers.

The Colleague Who Reports a Different Number

Two practitioners reporting different figures for the same thing is a normal state and a permanent source of organisational friction.

The differences are usually traceable to three causes. Different populations, where one used all customers and the other used active ones. Different windows, where one used calendar months and the other used rolling thirty days. Different treatments of an event, where one counted a cancellation at the date it was requested and the other at the date it took effect. Each is a legitimate choice and each produces a figure that is correct under its own definition.

The resolution is not to establish which figure is right. It is to publish the mapping: these two figures differ because of these three choices, and here is the reconciliation over a period where both can be computed. A reconciliation table converts a dispute into a known relationship, and it is the single most useful document two practitioners can produce together.

The second step is to agree which definition the organisation will use for which purpose, and to write it where the code lives. This is rarely achievable in one conversation and it is achievable over a quarter if both practitioners want it. Where one of them does not, the mapping alone still removes most of the friction, because the argument can be redirected to the definitions rather than conducted as a dispute about competence.

The third is to expect recurrence. New practitioners join, new definitions appear, the mapping ages. A quarterly reconciliation, performed by the people

who own the figures, keeps the relationship current and prevents a drift that nobody notices until two board papers disagree.

The Difference Between Evidence and Argument

Evidence and argument look similar in a meeting and behave completely differently under pressure, and practitioners are frequently asked to supply the second while believing they are producing the first.

An argument selects what supports a position. It is coherent, it is persuasive, and it is structured around a conclusion. Evidence is a body of observations assembled to be examined, and its structure is around the question rather than the answer. The same charts can serve either purpose, and the difference is not in the artefacts but in what would have been included had the data pointed elsewhere.

The practical test is the counterfactual inclusion. Would this analysis have contained the same elements if the result had been reversed. Where the answer is yes, the work is evidence. Where it is no, the work is an argument that has borrowed the appearance of evidence, and it will not survive a knowledgeable challenge because the missing elements are exactly what a challenger will find.

The second test is whether the strongest case against the conclusion is present in the document. A practitioner who has to be asked for it has produced a brief rather than an assessment. One who volunteers it has produced something an opponent can read without finding a gap, which is a far more durable form of persuasion.

The trade's reputation problem comes largely from this confusion. Practitioners are hired to produce evidence, are rewarded for arguments that support decisions, and discover that the second is easier and better paid. The discipline is to produce the first and to be explicit about the difference when a decision maker asks for the second.

The Subgroup That Was Found Afterwards

Subgroup analysis is where honest practitioners most often produce misleading results, because the search is natural and the reporting is selective.

An analysis shows no overall effect. The practitioner looks at segments and finds one where the effect is strong. This is exactly what a good practitioner

should do, because interventions do differ by segment, and it is also the mechanism by which a null result becomes a claim that an intervention works for young customers in the north.

The arithmetic is unforgiving. Ten segments examined at a conventional threshold will produce, under a true null, roughly a forty per cent chance of at least one apparently significant result. The finding is a property of the search and it is reported as a property of the segment, and the report rarely mentions how many segments were examined.

The discipline that holds is pre-specification plus labelling. Where the segments were named before the analysis, the finding is interpretable. Where they were discovered, it is a hypothesis and should be reported as one, with the number of comparisons stated. That sentence, this was one of ten segments examined and should be treated as a hypothesis, costs almost nothing and converts a likely false positive into a testable suggestion.

The second discipline is replication. A subgroup finding that matters should be tested in a fresh period or a fresh population before it influences anything expensive. This is rarely done because the finding is already persuasive, which is precisely why it needs to be.

The Prior That Was Really a Preference

Every estimate contains a prior, whether or not it is stated, and an unstated prior is a preference that has been smuggled into arithmetic.

The prior appears formally in Bayesian work and informally everywhere else. It is the belief about what the answer probably is before the data is examined, and it enters through the choice of method, the choice of population, the trimming of outliers, the decision about how many periods to include. A practitioner who trims the two most extreme months has encoded a belief that those months are not part of the world.

The discipline is to state the direction of the prior rather than to eliminate it. I expect this effect to be small because similar changes have been small, and I have chosen an approach that would detect a small one. That sentence makes the expectation examinable, and it also exposes the case in which the analysis could only have found what the practitioner expected.

The second discipline is to test the sensitivity of the conclusion to the prior. Where the answer flips under a reasonable alternative, the estimate is not robust and the practitioner knows it before somebody else discovers it. This is

a fifteen-minute check that changes the framing of every analysis in which it is performed.

The third is to notice when the prior is doing all the work. Where the data is thin and the conclusion is strong, the strength came from somewhere other than the data, and that somewhere is a belief that should be named. Naming it is not an admission of weakness; it is the difference between an estimate that can be examined and one that must be believed.

The Thing That Is Actually Kept

What the trade keeps is not the model. Models are replaced, and the replacement is usually better and rarely important.

What is kept is the habit of knowing, for any figure in the organisation, whether it is a measurement, an estimate or a hunch, and what conditions would need to hold for it to come right. An organisation that has that habit makes better decisions with worse models than one that has better models and no habit, because the first can tell the difference between a number it knows and a number it likes.

That habit lives in people rather than systems. It cannot be installed by a platform, a reorganisation kills it unless somebody put it in writing, and it is the only durable output of the trade. The chapters that follow are about how it is built, defended, and handed on.

ABOUT THE AUTHOR

Aiman JoeHanz writes the Building Tomorrow's People series. The data scientist keeps the estimate, and the estimate is a claim about a future nobody has seen. The book takes the question, the feature, the split that leaked, the metric that improved while the business did not, the model card written after the launch, and the conversation in which a confident number is asked to admit its range. It treats statistics as a moral discipline rather than a toolkit.