InquiryForge — Flashcards

A Spark & Anvil printable flashcard deck. Quiz yourself on the InquiryForge cast's curriculum — one card at a time.

How to use these cards

Fold each sheet down the middle line so the answers are hidden. Read the **question** on the left, say your answer out loud, then **unfold** to check the right side. Testing yourself — then checking — is one of the most powerful ways to remember. Get one wrong? That's the best kind of practice: try it again tomorrow.

Want real flashcards? Cut along the fold line and the rows into cards — question on one side, answer on the other.

---

1. Testable hypotheses; if-then

✂ Question (fold →)	Answer
Before designing an experiment, the FIRST step is to:	**state a clear, testable question or hypothesis** — A clear testable question guides the entire design.
A 'variable' in an investigation is:	**any factor that can change or be measured** — A variable is a factor that varies or is measured.
A hypothesis that is TOO vague to test can usually be fixed by:	**specifying the variable to change and the outcome to measure** — Naming the IV and a measurable DV converts a vague claim into a testable one.
Which is the BEST testable hypothesis?	**If plants get fertilizer, then they will grow taller than plants without it** — It is an if-then with a clear change and a measurable comparison.
The 'if' part of an if-then hypothesis usually names the:	**independent variable (what you change)** — The 'if' clause states what you deliberately change — the independent variable.
'Why is the sky beautiful?' is NOT a scientific question because it:	**asks about a value/opinion, not measurable evidence** — Beauty is a value judgment, not something evidence can settle.
Rewriting 'Are video games bad?' scientifically requires you to:	**define 'bad' as a measurable outcome (e.g. hours slept, test scores)** — You must operationalize 'bad' into a measurable outcome before it's testable.
'Does the color of a room affect how calm people feel?' becomes testable once you:	**operationalize 'calm' (e.g. measured heart rate or a rating scale)** — Defining 'calm' as a measurable outcome makes the question testable.
The 'then' part usually names the:	**dependent variable (what you measure)** — The 'then' clause states the outcome you measure — the dependent variable.
A scientifically TESTABLE question is one that:	**can be investigated with evidence you could actually collect** — A testable question can be answered by collecting evidence and could be shown wrong.
The value of a falsifiable hypothesis is that:	**a test can actually settle whether it holds** — Only claims that could be proven wrong can be genuinely tested.
A hypothesis is best stated as:	**an if-then prediction** — A hypothesis predicts: if I change X, then Y will happen.
A question you can answer just by looking it up is:	**not an experiment (it's research), though it can be valid inquiry** — Fact-lookup is legitimate research but isn't an experiment testing a variable.
A hypothesis and its prediction should be written:	**before collecting the data** — Stating the hypothesis first prevents fitting the story to the data after the fact.
A good scientific question is specific rather than broad because:	**specific questions can actually be tested and measured** — Specific questions define measurable variables you can investigate.
A testable hypothesis about salt and boiling water might be:	**Adding salt raises the temperature at which water boils** — It names a change (salt) and a measurable outcome (boiling temperature).
'Plants like nice music' is a weak hypothesis mainly because:	**'like' and 'nice' aren't measurable** — Vague, unmeasurable terms make the claim untestable until operationalized.
A question about what SHOULD	**an ethics/values question, not a testable empirical one** — 'Ought'

happen (an 'ought') is:	questions are about values; science tests 'is' questions with evidence.
Turning 'Does exercise help sleep?' into a testable form gives:	**Does 30 minutes of jogging before bed increase hours slept that night?** — The specific, measurable version names the change (jogging) and the measured outcome (hours slept).
Which is a TESTABLE question?	**Does more sunlight make bean plants grow taller?** — The sunlight-vs-height question can be measured and tested; the others are opinions/values.
A hypothesis must be FALSIFIABLE, meaning:	**it is possible to collect evidence that would prove it wrong** — Falsifiability means the claim could be contradicted by evidence — essential for science.
The reason scientists insist questions be testable is that:	**untestable claims can't be checked against reality** — Science advances only on claims evidence can check; untestable ones can't be evaluated.
A prediction differs from a hypothesis in that a prediction is:	**the specific expected result if the hypothesis is true** — A prediction states the concrete expected outcome derived from the hypothesis.
The synthesis of this kit is:	**science starts with a specific, measurable, falsifiable question** — A specific, measurable, falsifiable question is the foundation of every investigation.
A testable question makes an investigation POSSIBLE because it:	**tells you exactly what to change and what to measure** — It defines the variables, which the whole design then addresses.

2. Independent / dependent / controlled

✂ Question (fold →)	Answer
'Amount of sleep' as an independent variable would be tested by:	**giving groups different amounts of sleep and measuring an outcome** — To make sleep the IV you set different sleep amounts across groups.
The number of controlled variables in a real experiment is usually:	**many (all the factors you hold constant)** — Real experiments hold many factors constant to isolate the IV.
In 'Does fertilizer amount affect plant height?', the independent variable is:	**the amount of fertilizer** — Fertilizer amount is what's deliberately varied.
The synthesis of this kit is:	**a fair test changes one IV, measures the DV, and holds all else constant** — Isolating one IV, measuring the DV, and controlling the rest is the heart of a fair test.
Why keep controlled variables constant?	**so any change in the outcome can be attributed to the IV** — Holding other factors constant isolates the IV's effect on the DV.
A 'control GROUP' differs from a 'controlled VARIABLE': a control group is:	**a group that gets no treatment (a baseline)** — A control group is a no-treatment baseline; a controlled variable is a held-constant factor.
The INDEPENDENT variable is the one that:	**you deliberately change** — The independent variable is what the experimenter changes on purpose.
Labeling variables correctly matters because it:	**clarifies what causes what and keeps the design fair** — Correct variable roles make the causal logic and fairness of the design explicit.
A study measuring 'reaction time' after coffee has reaction time as its:	**dependent variable** — Reaction time is the measured outcome — the DV.
In the same study, the dependent variable is:	**the plant height** — Plant height is the measured outcome.
In the same study, a controlled variable would be:	**giving every plant the same amount of water** — Keeping water equal across plants controls that factor.
The dependent variable 'depends on' the:	**independent variable** — The DV's value may depend on the IV you manipulate.
The best way to describe the three variable roles is:	**change the IV, measure the DV, hold the controls constant** — Change IV → measure DV → hold controls constant is the core design logic.
Measuring the DV the SAME way for every group matters because:	**inconsistent measurement adds error and bias** — Consistent measurement keeps comparisons fair and reduces error.
In 'Does study time affect test scores?', 'test score' is the:	**dependent variable** — Test score is the measured outcome — the DV.
In that study, a controlled variable might be:	**the difficulty of the test given to everyone** — Giving everyone the same test difficulty controls that factor.
An experiment should ideally change:	**only ONE independent variable at a time** — Changing one IV at a time lets you attribute the effect to it.

CONTROLLED variables (constants) are factors you:	**keep the same for all groups** — Controlled variables are held constant so they can't affect the outcome.
Which is the DEPENDENT variable in 'Does louder music make people eat faster?'	**how fast people eat** — Eating speed is the measured outcome — the DV.
In a well-designed experiment, the only thing that should systematically differ between groups is:	**the independent variable** — Ideally groups differ only in the IV, so the DV difference is attributable to it.
If everything is held constant EXCEPT the IV, a difference in the DV is:	**evidence the IV had an effect** — With only the IV differing, a DV difference supports the IV's effect.
In 'Does room temperature affect typing speed?', room temperature is the:	**independent variable** — Room temperature is what's deliberately varied — the IV.
If you change TWO variables at once and see an effect, you:	**can't tell which variable caused it** — Changing two things leaves the cause ambiguous — a confounded design.
A variable you FORGOT to control that affects the result is a:	**confounding variable** — An uncontrolled influential factor is a confound.
The DEPENDENT variable is the one that:	**you measure as the outcome** — The dependent variable is the measured response that may depend on the IV.

3. Change one thing, hold the rest

✂ Question (fold →)	Answer
Giving the control group a sugar pill instead of the real drug is using a:	**placebo** — A sugar pill is a placebo controlling for expectation effects.
A controlled experiment's conclusion is trustworthy mainly because:	**the design rules out alternative explanations** — Good controls eliminate rival explanations, strengthening the causal conclusion.
The 'treatment group' is the one that:	**receives the change (the independent variable)** — The treatment group receives the IV manipulation.
A study testing a new fertilizer should give the control group:	**the same care but no fertilizer** — The control gets identical conditions minus the treatment.
The treatment and control groups should be as SIMILAR as possible at the start so that:	**any later difference is due to the treatment** — Similar starting groups mean post-treatment differences reflect the treatment.
Comparing treatment vs control lets you:	**see the effect of the treatment against a no-treatment baseline** — The difference between treatment and control isolates the treatment's effect.
A controlled experiment supports CAUSAL claims because:	**it isolates the IV as the only systematic difference** — Isolating the IV lets you attribute a DV difference to it — supporting causation.
If a result appears only because subjects EXPECTED it, you needed:	**a placebo control (and blinding)** — Expectation effects require placebo controls and blinding to rule out.
The main advantage of an experiment over just observing is:	**you can establish cause and effect** — Manipulating the IV under control lets you infer causation, not just correlation.
The number of subjects per group matters because:	**more subjects give more reliable results (less chance variation)** — Larger groups reduce the role of random variation, making effects clearer.
The strongest everyday example of a controlled experiment is:	**a clinical drug trial with treatment, placebo, and random assignment** — Randomized controlled trials with placebo controls are the gold standard.
The synthesis of this kit is:	**control + comparison + isolating one variable is how experiments reveal causes** — Controlled comparison isolating one IV is what lets experiments establish cause and effect.
If the treatment and control groups differ in MORE than the IV, the experiment is:	**confounded** — Extra differences confound the result, obscuring the IV's true effect.
Without a control group, you cannot tell whether:	**the outcome would have happened anyway** — A baseline is needed to know the change was due to the treatment, not something else.
The reason we RANDOMLY assign subjects to groups is to:	**make the groups similar in all other ways before treatment** — Random assignment balances unknown factors across groups so only the IV differs.
If plants in the treatment group also got more sunlight than the control, sunlight is a:	**confound (uncontrolled difference)** — The extra sunlight is an uncontrolled difference that confounds the fertilizer test.
	control for the psychological effect of receiving 'a treatment' — A

A PLACEBO is used to:	placebo (fake treatment) controls for expectation effects.
A control group given a placebo helps rule out:	**the placebo (expectation) effect** — A placebo control isolates the treatment's real effect from expectation.
'Double-blind' means:	**neither the subjects nor the researchers know who got the treatment** — Double-blind hides assignment from both sides, removing subject AND experimenter bias.
A 'variable' that changes together with the treatment and also affects the outcome must be:	**controlled or randomized away** — Such a confound must be controlled or balanced by randomization.
A controlled experiment is one where:	**you change one variable and keep the others constant to test its effect** — A controlled experiment isolates one variable's effect by holding the rest constant.
'Blinding' subjects means they:	**don't know whether they got the treatment or placebo** — Blinding hides group assignment to prevent expectation bias.
The 'control group' provides:	**a baseline to compare the treatment group against** — The control group (no treatment) is the comparison baseline.
A fair test means the groups are treated:	**identically except for the independent variable** — Fairness = identical treatment apart from the IV.
A well-controlled experiment holds constant everything EXCEPT:	**the independent variable** — Only the IV should differ systematically between groups.

4. The hidden third variable

✂ Question (fold →)	Answer
A study claiming 'people who meditate are happier' is confounded if:	**happier people are also more likely to choose to meditate** — Self-selection (happier people meditate) is a confound — the causal direction is unclear.
If you can't randomize (e.g. studying smokers), you should at least:	**measure and statistically adjust for likely confounds** — Observational studies measure and adjust for confounds, though causation stays weaker.
Confounds threaten the CONCLUSION that:	**the independent variable caused the outcome** — Confounds undermine the causal claim by offering an alternative explanation.
Two groups differ in the treatment AND in average age. To fix this you could:	**randomize assignment or match the groups on age** — Randomizing or matching on age removes it as a confound.
Older students score higher AND have studied longer. In a study-time study, AGE could be a:	**confound** — Age is linked to both study time and scores — a potential confound.
A confound can make it LOOK like:	**the treatment caused an effect it didn't** — Confounds create spurious apparent effects.
A 'lurking variable' is another name for:	**a confounding variable** — 'Lurking variable' is a synonym for a confound.
Ice-cream sales and drowning both rise in summer. The confound is:	**hot weather (drives both)** — Hot weather increases both — a confound, not a causal link between the two.
If people who take vitamins are also healthier eaters, diet is a:	**confound in a vitamin-vs-health study** — Diet is linked to both vitamin-taking and health — a confound.
A confound is different from measurement ERROR because a confound:	**systematically biases the comparison, not just adds random noise** — Confounds bias systematically; random error just adds noise.
A confounding variable is one that:	**is linked to both the treatment and the outcome, muddying the result** — A confound relates to both the IV and DV, making it unclear what caused the result.
A study finds coffee drinkers have more heart disease, but they also smoke more. Smoking is:	**a confound** — Smoking is linked to both coffee drinking and heart disease — a classic confound.
A scientist who says 'we controlled for age and diet' is:	**removing those as possible confounds** — 'Controlling for' a factor removes it as a confounding explanation.
The reason confounds don't ruin RANDOMIZED experiments is that:	**randomization spreads confounds evenly across groups** — Random assignment balances even unknown confounds across groups.
A confounding variable is a threat to a study's:	**internal validity (whether the IV really caused the DV)** — Confounds threaten internal validity — the causal link's trustworthiness.
A confounded study can still show	**a solid cause-and-effect conclusion** — Confounds allow correlation but

a correlation, but NOT:	block confident causal claims.
Confounds are most dangerous in:	**observational studies (no random assignment)** — Without random assignment, groups can differ in hidden ways — confounds abound.
Blocking (grouping similar subjects) helps with confounds by:	**controlling a known confound before randomizing within blocks** — Blocking holds a known confounder constant within groups, then randomizes.
The synthesis of this kit is:	**a confound is a hidden alternative cause — randomization and controls defeat it** — Spotting and eliminating hidden alternative causes (via randomization/controls) is essential to valid inference.
The whole reason confounds matter is that they:	**offer an alternative explanation for the results** — Confounds provide rival explanations that a good design must rule out.
The safest way to describe a confound to a friend is:	**a hidden third thing that could be the real cause** — A confound is a hidden alternative cause.
The BEST way to eliminate confounds in an experiment is:	**random assignment to groups** — Randomization balances confounds across groups so they don't bias the comparison.
Students who eat breakfast score higher, but they also sleep more. Sleep is a possible:	**confound** — Sleep is linked to both breakfast habits and scores — a possible confound.
To CHECK for a confound, ask whether some third factor:	**differs between groups AND could affect the outcome** — A confound both differs across groups and influences the outcome.
The presence of a plausible confound means a claim of causation should be:	**treated with caution and tested with a controlled experiment** — A plausible confound calls for skepticism and a controlled test.

5. Turn a fuzzy idea into a measurable one

✂ Question (fold →)	Answer
Using DIFFERENT measurement methods for different groups introduces:	**bias/error into the comparison** — Inconsistent measurement biases comparisons; use one method for all.
To compare 'intelligence' across a study you must first:	**operationalize it (e.g. a defined test score)** — A defined test score operationalizes the abstract idea for comparison.
Before comparing groups on 'stress,' a researcher must:	**operationalize stress into a measurable indicator** — A measurable stress indicator (e.g. cortisol, a validated scale) is required first.
Operationalizing lets a hypothesis become:	**testable with concrete data** — Concrete measures make the hypothesis empirically testable.
A vague outcome like 'did better' should be replaced with:	**a specific metric (e.g. test score out of 100)** — A concrete metric replaces vague judgments with measurable data.
A 'valid' measure is one that:	**actually measures what you intend** — Validity = the measure captures the intended concept.
If a survey question is ambiguous, the data it produces will be:	**hard to interpret and less valid** — Ambiguous items yield noisy, low-validity data.
A measure can be reliable but NOT valid if it:	**consistently measures the wrong thing** — A consistently-off measure is reliable (consistent) but invalid (wrong target).
The purpose of operational definitions is to make studies:	**replicable and comparable** — Clear measures let others repeat and compare the study.
A 'proxy' measure is:	**a measurable stand-in for something hard to measure directly** — A proxy is an indirect but measurable stand-in (e.g. income as a proxy for wealth).
Operationalizing 'happiness' might use:	**a validated self-report rating scale** — A validated rating scale gives a measurable, comparable value for happiness.
Operational definition of 'exercise intensity' might be:	**heart rate as a percentage of maximum** — % of max heart rate is a measurable operational definition of intensity.
Operationalizing is essentially the bridge between:	**an abstract idea and measurable data** — It links the concept to concrete, collectable measurements.
The synthesis of this kit is:	**you can't test what you can't measure — operationalize first** — Turning fuzzy ideas into measurable definitions is a prerequisite for real inquiry.
'Healthy' is hard to measure until you operationalize it as, e.g.:	**resting heart rate or BMI** — Choosing a measurable proxy (heart rate, BMI) operationalizes 'healthy'.
Operationalizing early in a study helps because it:	**prevents shifting definitions after seeing data** — Defining measures up front stops post-hoc redefinition that biases conclusions.
A risk of proxy measures is that:	**the proxy may not perfectly capture the real concept** — Proxies can diverge from the true concept, weakening validity.
To 'operationalize' a variable means	**define it as something you can actually measure** — Operationalizing

to:	turns an abstract idea into a concrete, measurable definition.
A good operational definition is:	**specific enough that two people would measure it the same way** — Consistent, reproducible measurement is the mark of a good operational definition.
Measuring 'reaction time' with a stopwatch by hand vs a computer differs in:	**precision (the computer is more precise)** — The tool affects measurement precision; a computer times more precisely.
The BEST operational definition of 'screen time' is:	**hours per day of device use, logged automatically** — Automatically logged hours/day is precise, reproducible, and valid.
A 'reliable' measure is one that:	**gives consistent results when repeated** — Reliability = consistency across repeated measurements.
Measuring 'plant growth' operationally could be:	**change in height in cm over two weeks** — Height change in cm over a set time is a precise operational measure.
Two researchers measuring the same thing and agreeing shows:	**good inter-rater reliability** — Agreement between raters indicates the measure is reliably defined.
A study that never defines its key term precisely is vulnerable to:	**the criticism that it measured something else** — Without a clear definition, critics can argue a different thing was measured.

6. Representative samples; selection bias

✂ Question (fold →)	Answer
'Convenience sampling' (whoever is easy to reach) risks:	**an unrepresentative, biased sample** — Convenience samples over-represent the easy-to-reach, biasing results.
Polling only your friends about a school issue is an example of:	**convenience sampling (biased)** — Friends are an easy, unrepresentative group — convenience sampling bias.
A famous polling failure occurred when the sample:	**was drawn from a biased list (e.g. phone/car owners)** — Biased sampling frames (like early phone directories) produced famously wrong predictions.
A biased sample threatens a study's:	**external validity (generalizability)** — Sampling bias undermines generalizing to the population — external validity.
If a sample under-represents a key group, conclusions about that group are:	**unreliable / not generalizable** — Missing groups can't be validly described from the sample.
Random ASSIGNMENT (to groups) supports:	**cause-and-effect conclusions** — Random assignment balances confounds, supporting causal claims (distinct from selection).
The gold standard for a generalizable survey is:	**a random sample from a complete, unbiased frame** — Random selection from an unbiased frame is the standard for generalizable surveys.
Surveying only people at a gym about exercise habits causes:	**selection bias (gym-goers exercise more)** — The venue over-selects exercisers, biasing the sample.
The population is:	**the entire group you want to draw conclusions about** — The population is the whole group of interest.
The synthesis of this kit is:	**how you SELECT the sample determines whether results generalize** — Representative (usually random) selection is what makes generalization valid.
A 'representative' sample is one that:	**reflects the population's key characteristics** — A representative sample mirrors the population so results generalize.
Selection bias occurs when:	**some members of the population are systematically more likely to be chosen** — Selection bias skews the sample away from the population.
'Voluntary response' samples are biased because:	**people with strong opinions are more likely to respond** — Self-selected responders skew toward the strongly opinionated.
Increasing sample size when the METHOD is biased will:	**produce a precisely-wrong (still biased) estimate** — A bigger biased sample is just more confidently wrong.
The main reason to sample RANDOMLY is to:	**avoid systematic bias and allow generalization** — Random selection avoids systematic bias so the sample generalizes to the population.
A larger sample mainly reduces:	**random sampling error (not bias)** — Size reduces random error but a biased sampling method stays biased no matter the size.
A stratified sample:	**divides the population into groups and samples each** — Stratified sampling ensures each subgroup is represented.
The safest interpretation of a	**it may not generalize beyond that convenient group** — Convenience

convenience-sample result is:	results are limited to the sampled group unless shown representative.
Non-response bias occurs when:	**those who don't respond differ from those who do** — If non-responders differ systematically, the responding sample is biased.
A 'sample' is:	**a subset of the population you actually study** — A sample is the portion of the population you measure.
'Random SELECTION' (sampling) supports:	**generalizing results to the population** — Random selection is what lets you generalize the sample's results to the population.
Asking 'do you support the popular new policy?' biases responses via:	**the wording (leading question)** — Leading wording ('popular') biases answers — a response bias, separate from sampling.
To sample a school fairly, you might:	**randomly select students from a full roster** — A random draw from the full roster avoids selection bias.
A sampling FRAME is:	**the actual list from which the sample is drawn** — The frame is the source list; a biased frame biases the sample.
A 'random sample' means:	**every member has an equal chance of being chosen** — Random sampling gives each member an equal selection chance, reducing bias.

7. Randomize; control groups

✂ Question (fold →)	Answer
Randomization addresses confounds; blinding addresses:	**expectation/observer bias** — Randomizing balances confounds; blinding removes expectation and observer bias.
A 'block' in a block design is:	**a group of subjects similar on a known variable** — A block groups subjects alike on a known influential variable.
The purpose of a fair comparison is ultimately to:	**attribute any outcome difference to the treatment** — A fair comparison lets you credit the treatment for the difference.
A study without random assignment can still show correlation but should avoid claiming:	**causation** — Without random assignment, causal claims are unsupported.
Randomized controlled trials (RCTs) are considered strong evidence because:	**randomization + control isolate the treatment's causal effect** — RCTs combine randomization and controls to yield strong causal evidence.
If you can't randomize (ethics/practicality), the study becomes:	**observational (weaker for causation)** — Without randomization it's observational, and causal claims weaken.
A randomized BLOCK design first:	**groups similar subjects, then randomizes within each block** — Blocking controls a known confounder by grouping, then randomizes within blocks.
Randomization is typically done using:	**a random number generator or equivalent chance device** — True randomization uses a chance mechanism, not human choice.
Comparing a treatment group to NO control group is flawed because:	**you don't know what would have happened anyway** — Without a control you lack the counterfactual baseline.
Random assignment makes the treatment and control groups:	**probabilistically equivalent at baseline** — Randomization makes groups equivalent on average before treatment.
If a doctor assigns the sickest patients to the new drug, the comparison is:	**biased (not randomized)** — Assigning by health status confounds the comparison — not randomized.
The reason NOT to let subjects choose their own group is:	**self-selection creates confounds** — Letting subjects choose introduces systematic differences — confounds.
A matched-pairs design:	**pairs similar subjects (or uses each subject twice) then randomizes within pairs** — Matched pairs compares within closely-matched units, sharpening the treatment comparison.
The fairest way to decide who gets the treatment is:	**a random draw** — A random draw removes human bias from assignment.
The key benefit of random assignment is that it:	**balances known AND unknown confounds across groups** — Randomization spreads even unknown confounders evenly, enabling causal claims.
Random ASSIGNMENT to treatment	**make the groups similar in all other ways before treatment** —

and control groups is used to:	Randomizing who gets the treatment balances confounds across groups.
Randomization defeats confounds because it:	**prevents any systematic difference between groups except the treatment** — By randomizing, no factor is systematically tied to a group except the assigned treatment.
Blocking is useful when:	**a known variable strongly affects the outcome** — Blocking reduces variability from a known influential factor.
A control group in a randomized experiment provides:	**the counterfactual: what happens without the treatment** — The randomized control estimates what would have happened without treatment.
Flipping a coin to decide who gets the new study method is a form of:	**random assignment** — A coin flip randomly assigns subjects to groups.
A 'fair comparison' between groups requires:	**the groups be alike except for the treatment** — Comparable groups differing only in treatment make the comparison fair.
Random ASSIGNMENT supports causation; random SELECTION supports:	**generalization to the population** — Assignment → causation; selection → generalization. Different jobs.
The synthesis of this kit is:	**random assignment + a control group is how experiments make fair, causal comparisons** — Randomization plus a control group underpins fair comparison and causal inference.
A completely randomized design:	**assigns all subjects to treatments purely at random** — In a completely randomized design, every subject is randomly assigned with no restrictions.
If groups end up different by CHANCE despite randomization, this is:	**reduced by larger samples** — Random imbalance can happen but shrinks as sample size grows.

8. A link is not a cause

✂ Question (fold →)	Answer
The key warning of this topic is:	**correlation does not prove causation** — Two variables can correlate without one causing the other.
Countries with more TVs have longer lifespans. The likely explanation is:	**wealth (a third variable) drives both** — Wealth causes both more TVs and better healthcare — a spurious correlation.
'Correlation' means two variables:	**tend to change together** — Correlation is a statistical association — they move together.
The synthesis of this kit is:	**a link is a clue, not a verdict — establish causation with controlled tests** — Correlation motivates investigation; controlled experiments establish causation.
A correlation of zero means:	**no linear association was detected** — Zero correlation indicates no linear relationship in the data.
If randomly assigning X changes Y, you have evidence that:	**X causes Y** — A randomized manipulation of X changing Y is evidence X causes Y.
Ice cream sales correlate with drownings because:	**a third variable (hot weather) drives both** — A common cause (heat) creates the correlation — not a direct causal link.
Observational studies can find correlations but struggle to:	**rule out confounds and establish causation** — Without randomization, confounds make causal claims unreliable.
Even a genuine cause can show a WEAK correlation if:	**other factors (noise/confounds) also affect the outcome** — Real causes can be masked by noise and other influences, weakening correlation.
A strong correlation with a plausible mechanism still needs:	**experimental confirmation for a causal claim** — Even a plausible mechanism requires a controlled test to confirm causation.
To argue A causes B, the strongest single move is to:	**run a randomized experiment manipulating A** — Randomly manipulating A and observing B is the strongest causal evidence.
The phrase 'correlation implies causation' is:	**a common logical error to avoid** — Assuming causation from correlation is a classic fallacy.
Which claim is safe from a correlation alone?	**'X and Y are associated'** — Only association is justified by correlation; causal wording is not.
A 'spurious correlation' is one that:	**appears real but is due to a third variable or chance** — Spurious correlations arise from confounds or coincidence, not a true link.
The direction of causation is hardest to determine in:	**observational data** — Observational data often can't reveal which variable drives which.
A responsible headline about an observational finding would say:	**'linked to' rather than 'causes'** — 'Linked to' respects that observation shows association, not causation.
Two variables can be correlated purely by:	**coincidence in a small or cherry-picked dataset** — Chance can produce correlations, especially in small or selectively-chosen data.
Three explanations for a correlation between A and B	**A causes B, B causes A, or a third variable causes both** — A correlation could

are:	be A→B, B→A, or a common cause (or coincidence).
'Confounding' explains many spurious correlations because:	**a hidden variable causes both measured variables** — A confound driving both variables produces a misleading correlation.
A scatterplot showing points trending upward indicates:	**a positive correlation (not necessarily causation)** — An upward trend shows positive correlation; causation needs more.
The scientist's discipline about causation is to:	**withhold causal claims until a controlled test supports them** — Reserve causal language for when controlled evidence supports it.
To move from correlation to causation you generally need:	**a controlled experiment with random assignment** — Randomized experiments isolate the causal effect that correlation alone can't establish.
'Reverse causation' means:	**the effect might actually be causing the supposed cause** — Reverse causation: B→A when you assumed A→B.
More firefighters at a fire correlates with more damage. The real driver is:	**the fire's size (a third variable)** — Bigger fires cause both more firefighters and more damage — a third variable.
'Smoking causes cancer' became accepted because:	**many lines of evidence including experiments and mechanisms converged** — Convergent evidence (experiments, mechanisms, dose-response) established the causal claim.

9. Variability; is the difference real?

✂ Question (fold →)	Answer
'Cherry-picking' data means:	**selecting only the results that support your claim** — Cherry-picking reports favorable results and hides the rest — a distortion.
The danger of testing MANY comparisons is:	**some will look 'significant' by chance alone** — Testing many things inflates false positives — chance produces some 'hits'.
A tiny but statistically significant effect in a huge sample may be:	**real but too small to matter practically** — Huge samples can detect real but practically trivial effects.
'Noise' in data refers to:	**random variation not due to the variable of interest** — Noise is random fluctuation obscuring the real signal.
The honest reporting of a result includes:	**the uncertainty (e.g. a confidence interval or error bars)** — Reporting uncertainty lets others judge the signal against the noise.
An outlier is:	**a data point far from the rest** — An outlier lies far from the bulk of the data and can distort analysis.
A larger sample makes it easier to:	**detect a real signal above the noise** — More data averages out noise, revealing the signal.
Random noise tends to:	**average out over many observations** — Because it's random, noise cancels out as observations accumulate.
'Signal' refers to:	**the real underlying effect or pattern** — Signal is the genuine effect you're trying to detect.
To tell signal from noise, scientists use:	**statistics (e.g. significance tests)** — Statistical tests estimate whether a difference exceeds what chance would produce.
A result that disappears when the study is repeated was probably:	**noise, not a real signal** — Effects that don't replicate are likely chance/noise.
'Statistical significance' roughly means:	**the result is unlikely to be due to chance alone** — Significance indicates the result is unlikely under pure chance — not that it's large/important.
The best defense against being fooled by noise is:	**replication and adequate sample size** — Replicating with enough data separates real signals from noise.
Repeating measurements and averaging helps because it:	**reduces the influence of random noise** — Averaging cancels random error, sharpening the signal estimate.
If two group averages differ but their data hugely overlap, the difference is:	**likely not meaningful (swamped by noise)** — Large overlap suggests the difference could easily be noise.
A small difference between two groups might be:	**just noise (random variation), not a real effect** — Small differences can arise from chance alone — you must check if they're real.
A 'p-value' is small when:	**the observed result is unlikely if there were no real effect** — A small p-value means the data would be unlikely under the no-effect assumption.
A confidence interval expresses:	**a range of plausible values for the true effect** — A confidence interval shows the uncertainty range around an estimate.

The synthesis of this kit is:	**real conclusions require separating the signal from random noise** — Distinguishing genuine effects from chance variation is central to honest data analysis.
Before concluding a difference is real, you should ask:	**could random variation alone produce this?** — The core question is whether chance alone could explain the observed difference.
'Variability' in a dataset is measured by things like:	**the spread (e.g. standard deviation)** — Spread statistics (like standard deviation) quantify variability/noise.
A wide confidence interval indicates:	**a lot of uncertainty (often from small samples/high noise)** — Wide intervals mean the estimate is imprecise — much noise or little data.
If a result vanishes with a much larger sample, the original was likely:	**noise (a chance fluctuation)** — An effect that disappears with more data was probably random noise.
Statistical significance is NOT the same as:	**practical importance** — A significant effect can still be tiny/unimportant in practice.
Two honest ways to increase signal-to-noise are:	**bigger samples and more precise measurement** — More data and better measurement both raise the signal-to-noise ratio.

10. Axes, scales, misleading charts

✂ Question (fold →)	Answer
A truthful bar chart usually has a y-axis that:	**starts at zero** — For bar charts, a zero baseline keeps bar lengths proportional to values.
A logarithmic scale is appropriate when:	**values span many orders of magnitude** — Log scales suit data spanning huge ranges; misused, they can mislead.
A chart is a form of ARGUMENT, so you should:	**evaluate it as critically as any claim** — Charts make arguments; read them with the same skepticism as claims.
A 'gee-whiz' graph exaggerates a trend mainly through:	**axis scaling tricks** — Gee-whiz graphs use scaling tricks (like cropped axes) to overstate trends.
To judge a trend fairly, you should:	**look at the actual axis scale, not just the line's steepness** — The scale determines what a slope really means; steepness alone can mislead.
A pictograph using bigger IMAGES for bigger values can mislead because:	**area/volume grows faster than the value** — Scaling an image's size makes area grow faster than the number, exaggerating differences.
Cherry-picking the TIME RANGE of a graph can:	**hide an overall trend by showing only a favorable window** — Selecting a convenient window can conceal the broader trend.
The honest question to ask of any chart is:	**does the visual accurately represent the underlying numbers?** — A fair chart's visual matches the real data proportions.
The FIRST thing to check on any graph is:	**the axis labels and units** — Axis labels and units tell you what's actually being shown.
Smoothing or averaging data on a graph can:	**hide real short-term variation** — Over-smoothing can conceal genuine variability.
A common way to EXAGGERATE a difference is to:	**start the y-axis somewhere other than zero** — A truncated (non-zero) y-axis visually magnifies small differences.
If two lines are on different scales, comparing their heights directly is:	**invalid** — Different scales make direct height comparison meaningless.
A correlation shown on a scatterplot still does NOT prove:	**causation** — A scatterplot correlation is not causal evidence.
A 3-D pie chart can distort perception by:	**making nearer slices look larger** — 3-D perspective distorts slice sizes, biasing perception.
Reordering categories on a bar chart to tell a story is:	**potentially misleading if it implies a false pattern** — Strategic ordering can suggest a pattern that isn't in the data.
A graph with NO axis labels is:	**untrustworthy — you can't tell what it shows** — Without labels/units you can't interpret or trust the graph.
Two honestly-made graphs of the same data should:	**lead to the same conclusion** — Honest visualizations of the same data agree in their message.
A graph that omits the data SOURCE should be treated with:	**caution (unverifiable)** — Without a source you can't verify the data — be cautious.

A dual-axis chart can mislead by:	**implying a relationship between unrelated scales** — Two different y-axes can be scaled to fake a relationship.
The synthesis of this kit is:	**graphs can inform or mislead — always read the axes, scale, and source** — Honest graph-reading means checking axes, scale, and source before trusting the visual.
Error bars on a graph show:	**the uncertainty or variability in the data** — Error bars convey uncertainty/variability around each value.
A line graph whose y-axis runs 98 to 100 will make a 1-point change look:	**dramatic** — A cropped 98–100 axis exaggerates a small change.
A bar chart with a y-axis starting at 90 (not 0) can make a tiny difference look:	**huge** — Cropping the baseline inflates the apparent size of differences.
The best defense against a misleading chart is to:	**look at the raw numbers and axis scale yourself** — Checking the underlying numbers and scale defeats visual tricks.
To read a graph critically, compare the claim to:	**what the axes and scale actually show** — Check whether the axes/scale support the stated claim.

11. Does it hold?

✂ Question (fold →)	Answer
Pre-registering a study (declaring the plan first) helps replication by:	**preventing after-the-fact analysis choices that inflate false positives** — Pre-registration locks the plan, curbing flexible analyses that produce fragile findings.
Replication is part of what makes science:	**self-correcting** — By re-testing findings, science corrects its errors over time.
'p-hacking' means:	**trying many analyses until one looks significant** — P-hacking searches analyses for a significant result — a source of non-replicable findings.
A study with a tiny sample that finds a huge effect should be:	**replicated before it's believed** — Small samples produce unstable estimates; replication is essential.
Sharing data and methods openly supports:	**others' ability to reproduce and replicate the work** — Open data/methods let others check and repeat the study.
Replication means:	**repeating a study to see if the result holds** — Replication re-runs a study to test whether the finding recurs.
One reason results fail to replicate is:	**the original was a false positive (chance)** — Some original findings are chance false positives that don't recur.
A 'conceptual replication' tests the same idea with:	**different methods or measures** — Conceptual replication checks the idea generalizes across methods.
Reproducing an ANALYSIS from shared data checks for:	**errors in the analysis (computational reproducibility)** — Re-running the analysis on the same data catches analysis errors.
Replication turns a single finding into:	**reliable knowledge** — Repeated confirmation is how findings become reliable knowledge.
The reason one surprising study makes headlines but shouldn't be trusted yet is:	**it hasn't been replicated** — Novel single findings need replication before confidence.
If a famous result fails many replications, the responsible response is to:	**doubt the original finding** — Repeated replication failures warrant doubting the original claim.
Replication protects against being fooled by:	**chance, error, and bias in a single study** — Repeating guards against one-off chance, mistakes, and bias.
'Reproducibility' (vs replication) often means:	**getting the same result from the SAME data and methods** — Reproducibility = same data/analysis yields the same result; replication = a fresh study.
A result that replicates is:	**more trustworthy** — Repeated confirmation increases confidence the effect is real.
A finding confirmed by three independent labs is:	**more trustworthy than a single lab's result** — Independent replication across labs strongly supports a real effect.
Publication bias (only publishing	**the literature overstates how often effects are real** — Hiding null results

positive results) harms replication because:	makes effects look more reliable than they are.
The 'replication crisis' refers to:	**many published findings failing to replicate** — In several fields, a surprising fraction of published effects didn't hold up on repetition.
The synthesis of this kit is:	**a result you can't repeat isn't knowledge yet — replication is the test of truth** — Reproducible, replicable findings are what science treats as reliable.
The scientist's stance toward their own exciting result should be:	**'let's see if it replicates'** — Genuine confidence comes from replication, even of your own findings.
A finding that replicates across DIFFERENT labs and samples is:	**especially robust** — Independent replications across contexts strongly support a real effect.
A single study, even a good one, should be treated as:	**preliminary until replicated** — One study is a starting point; confidence grows with replication.
A large, pre-registered replication is valuable because it:	**provides a strong, unbiased test of the effect** — Large pre-registered replications are among the strongest tests of a claimed effect.
A field with high replication rates is generally:	**more trustworthy** — High replication indicates rigorous, trustworthy findings.
A 'direct replication' repeats the study:	**as closely as possible to the original** — Direct replication mirrors the original procedure to test the same effect.

12. Support a claim with evidence

✂ Question (fold →)	Answer
Good evidence is:	**relevant, sufficient, and accurate** — Evidence must be on-point, enough, and correct to support a claim.
An anecdote (one story) is weak evidence because:	**it may not represent the general case** — A single story can be unrepresentative — weak support for a general claim.
Evidence quality matters more than:	**the number of sources alone** — Many weak sources don't beat solid, relevant evidence.
Reasoning often invokes:	**a scientific principle or mechanism** — Reasoning connects evidence to claim, usually via a principle or mechanism.
Counter-evidence should be:	**acknowledged and addressed, not hidden** — Honest arguments engage evidence that challenges the claim.
In CER, the REASONING is:	**the explanation of WHY the evidence supports the claim** — Reasoning links evidence to claim, often via a scientific principle.
A strong scientific argument includes:	**a claim, supporting evidence, AND reasoning** — All three (C, E, R) make a complete argument.
In CER, the EVIDENCE is:	**the data/observations that support the claim** — Evidence is the relevant data supporting the claim.
When evidence contradicts your claim, the scientific response is to:	**revise the claim** — Contradicting evidence should prompt revising the claim.
The point of CER in class is to:	**build the habit of supporting claims with evidence and reasoning** — CER trains evidence-based reasoning — a core scientific habit.
In CER, the CLAIM is:	**the answer or conclusion you are arguing for** — The claim is the position/conclusion being defended.
If someone gives a claim and data but no reasoning, ask:	**how does that evidence support the claim?** — Demand the reasoning that links evidence to claim.
A claim WITHOUT evidence is:	**just an assertion/opinion** — Without evidence, a claim is a bare assertion.
The reasoning should explain the evidence in terms of:	**the underlying science, making the link explicit** — Reasoning makes the evidence-to-claim connection explicit via the science.
A well-reasoned argument can still be wrong if:	**its evidence is inaccurate** — Even good reasoning fails if built on inaccurate evidence.
A claim like 'this drug works' needs evidence such as:	**results from controlled trials** — Controlled-trial results are the appropriate evidence for a 'works' claim.
'This soil is better because plants grew 5 cm taller in it' is missing the CER element of:	**reasoning (why the height shows 'better')** — It states a claim and evidence but not the reasoning linking them.
To evaluate someone's argument, first identify:	**their claim, evidence, and reasoning** — Separating C/E/R clarifies whether the argument holds.

An argument that only restates the claim as 'evidence' is:	**circular (no real support)** — Restating the claim as its own support is circular reasoning.
The synthesis of this kit is:	**a scientific claim is only as strong as the evidence and reasoning behind it** — Claims must be backed by sufficient, accurate evidence and explicit reasoning.
A claim that goes BEYOND the evidence is:	**overclaiming (unsupported)** — Stating more than the evidence supports is overclaiming.
Evidence that is 'sufficient' means:	**there is enough of it to support the claim** — Sufficiency = enough evidence to warrant the claim.
A complete CER response to 'which soil grew taller plants?' includes:	**the claim, the height data, and why that data supports it** — State the claim, cite the height data, and reason why it supports the claim.
The reasoning step is important because evidence:	**doesn't explain itself — you must connect it to the claim** — You must explain how/why the evidence supports the claim.
In science, disagreements are ideally settled by:	**evidence and reasoning, not authority** — Scientific disputes are resolved by evidence and argument, not status.

13. Construct + critique arguments

✂ Question (fold →)	Answer
A respectful challenge to a claim you agree with is:	**still valuable — it tests the claim** — Testing even agreeable claims strengthens them and guards against bias.
When two peers disagree, the resolution should come from:	**examining the evidence together** — Disagreements are resolved by jointly examining the evidence.
Consensus in science is built through:	**repeated argument, evidence, and replication over time** — Scientific consensus emerges from accumulated evidence and argument.
Argument-Driven Inquiry (ADI) centers on:	**constructing and critiquing evidence-based arguments** — ADI has students build and critically evaluate arguments from evidence.
The 'burden of proof' rests on:	**the person making the claim** — Whoever asserts a claim must support it with evidence.
The difference between debate and scientific argumentation is that argumentation aims to:	**reach the best-supported conclusion, not just to win** — Scientific argumentation seeks truth via evidence, not victory.
In a class argumentation session, the productive norm is:	**challenge ideas rigorously while respecting people** — Rigorous idea-critique plus interpersonal respect is the healthy norm.
Considering ALTERNATIVE explanations makes an argument:	**stronger (you've ruled out rivals)** — Addressing alternative explanations strengthens the case.
A 'rebuttal' in argumentation is:	**a response addressing counter-arguments** — A rebuttal answers objections and counter-evidence.
Changing your mind when the evidence warrants it is:	**a scientific strength, not a weakness** — Updating beliefs on evidence is core to good science.
A strong argument anticipates:	**the strongest objections and answers them** — Addressing the strongest counter-arguments makes a case robust.
An argument session ends well when:	**the group converges on the claim best supported by evidence** — Success is convergence on the evidence-best conclusion.
A claim supported by strong evidence but poor reasoning should be:	**improved by fixing the reasoning** — Strengthen the reasoning rather than discard a well-evidenced claim.
A hallmark of a good scientific arguer is:	**willingness to follow the evidence even against their preference** — Following evidence over preference is the mark of scientific integrity.
A productive critique focuses on:	**the evidence and reasoning, not the person** — Critique the argument's substance, never the individual.
A 'strawman' argument:	**misrepresents an opponent's position to attack it easily** — A strawman distorts the opponent's view into an easy target.
The goal of critiquing a peer's argument is to:	**improve the argument by testing its evidence and reasoning** — Critique strengthens arguments by probing evidence and reasoning respectfully.
The synthesis of this kit is:	**science is a social practice of building and critiquing evidence-based arguments** — Constructing and critiquing arguments from evidence is how

	scientific knowledge is forged.
A good question to ask of any argument is:	**'What evidence supports that, and is it enough?'** — Probing the evidence and its sufficiency tests the argument.
A logical fallacy weakens an argument by:	**using flawed reasoning instead of sound evidence** — Fallacies substitute flawed reasoning for genuine support.
An 'ad hominem' argument attacks:	**the person instead of their argument** — Ad hominem targets the person, dodging the actual argument.
Presenting your argument to peers for critique is valuable because:	**others spot weaknesses you missed** — Peer critique surfaces blind spots and strengthens the argument.
The reason argumentation is a scientific PRACTICE is that:	**science advances by publicly testing claims through argument** — Public argument and critique is how the community tests and refines claims.
Attacking a classmate's character instead of their evidence is a:	**logical fallacy (ad hominem)** — Targeting the person, not the argument, is the ad hominem fallacy.
If a peer's evidence is weak, the respectful move is to:	**point out the specific weakness and ask for more** — Name the specific gap and invite stronger evidence.

14. When you can't randomize

✂ Question (fold →)	Answer
The reason we can't always run experiments is:	**ethics, cost, or feasibility** — Ethical, financial, and practical limits sometimes rule out experiments.
A 'cohort study' follows a group over time and is:	**observational** — Cohort studies observe groups over time without manipulation — observational.
Studying smoking's effects on humans is usually observational because:	**it would be unethical to assign people to smoke** — You can't ethically assign people to smoke, so such studies are observational.
An EXPERIMENTAL study:	**deliberately manipulates a variable and can use random assignment** — Experiments manipulate the IV, ideally with random assignment.
A study that follows coffee drinkers and non-drinkers over years WITHOUT assigning coffee is:	**observational** — No assignment of the variable means it's observational.
The difference in causal strength between the two designs comes down to:	**whether the variable was manipulated with randomization** — Manipulation + randomization is what elevates experiments for causation.
Observational studies CAN establish:	**associations and generate hypotheses** — They reveal associations and suggest hypotheses to test experimentally.
The main advantage of experiments is that they:	**can establish cause and effect** — Manipulation + randomization lets experiments support causal claims.
Observational studies are used when experiments are:	**unethical or impractical** — When you can't or shouldn't manipulate the variable, you observe instead.
The key WEAKNESS of observational studies is:	**confounds make causal claims uncertain** — Without randomization, confounds cloud causal interpretation.
An OBSERVATIONAL study:	**measures variables without manipulating them** — Observational studies record variables as they naturally occur.
The presence of confounds is the core reason observational studies:	**can't easily claim causation** — Uncontrolled confounds are why observation struggles with causation.
A responsible scientist matches the STUDY DESIGN to:	**the question and what is ethical/feasible** — Design should fit the question and real-world constraints, not the wished-for result.
The BEST design to prove a new teaching method causes higher scores is:	**a randomized experiment assigning classes to methods** — Randomly assigning the method isolates its causal effect.
A 'natural experiment' occurs when:	**a naturally-occurring event mimics random assignment** — Natural experiments exploit chance-like real-world events to approximate randomization.
Combining many observational studies (meta-analysis) can:	**strengthen an association but still not prove causation alone** — Pooling observations strengthens evidence but shared confounds limit causal proof.

Observational studies are often more:	**realistic/representative of natural conditions** — By observing real settings, they can be more ecologically realistic.
A prospective study:	**follows subjects forward in time** — Prospective studies track subjects going forward — observational.
An observational finding that 'X is associated with Y' should be reported as:	**an association, pending experimental confirmation** — Report associations cautiously until experiments confirm causation.
The synthesis of this kit is:	**experiments establish causation; observational studies reveal associations when you can't experiment** — Choose experiments for causation when possible; use observation for associations otherwise.
To strengthen causal claims from observational data, researchers:	**measure and statistically adjust for confounds** — Measuring and adjusting for known confounds partially strengthens the inference.
A 'case-control study' compares those with and without an outcome and is:	**observational** — Case-control studies look back at existing groups — observational.
A retrospective study:	**looks back at existing data/records** — Retrospective studies analyze past records — observational.
The strongest evidence for causation typically comes from:	**randomized controlled experiments** — RCTs provide the strongest causal evidence.
If a treatment can be ethically and practically randomized, you should prefer:	**an experiment** — When feasible and ethical, experiments give stronger causal evidence.

15. Over-/under-claiming; limits

✂ Question (fold →)	Answer
'Absence of evidence' for an effect is:	**not the same as evidence of no effect** — Failing to find an effect isn't proof none exists (could be low power).
A conclusion that could apply to a CONTESTED real-world issue should:	**present the evidence and trade-offs, not a personal verdict** — On contested issues, lay out evidence and trade-offs; avoid a scored verdict (anti-evangelism).
Moving from result to conclusion, the discipline is to:	**claim exactly what the evidence supports — no more, no less** — Match the conclusion precisely to the evidence.
Interpreting a graph, your conclusion should match:	**what the axes and data actually show** — Conclusions must fit the actual data, not the dramatic impression.
The synthesis of this kit is:	**a good conclusion is calibrated to the evidence and honest about its limits** — Honest, appropriately-limited conclusions matched to the evidence are the goal.
Suggesting FUTURE studies in a conclusion is:	**good practice (science is ongoing)** — Noting next steps reflects science as an ongoing process.
If a result is not statistically significant, the conclusion should:	**not claim a real effect was found** — A non-significant result doesn't support claiming a real effect.
Generalizing beyond the STUDIED POPULATION is:	**unjustified without evidence it applies** — Conclusions apply to the studied population unless shown to extend.
A conclusion that ignores contradicting data is:	**biased/dishonest** — Omitting contradicting evidence biases the conclusion.
A conclusion should connect back to:	**the original question/hypothesis** — Good conclusions answer the question the study asked.
If results were correlational, the conclusion must avoid:	**claiming causation** — Correlational results don't license causal conclusions.
A study on 20 people should generalize to:	**cautiously, noting the small, specific sample** — Small, narrow samples limit how far you can generalize.
'Underclaiming' means:	**failing to state a conclusion the evidence actually supports** — Underclaiming ignores what the data does support.
'Overclaiming' means:	**stating a conclusion stronger than the evidence supports** — Overclaiming asserts more than the data warrants.
Every conclusion should state its:	**limitations** — Acknowledging limitations is part of an honest conclusion.
Reporting the EFFECT SIZE (not just significance) matters because:	**it tells you how big/important the effect is** — Effect size conveys practical importance beyond mere significance.
A conclusion should be:	**supported by the data and no broader than the evidence allows** — Conclusions must stay within what the evidence supports.
From a study of 15 local students, concluding 'all teenagers worldwide	**overclaiming** — Generalizing from 15 local students to all teens

behave this way' is:	worldwide overclaims.
A conclusion drawn from a biased sample should:	**note that it may not generalize** — Acknowledge that a biased sample limits generalization.
A conclusion can be weakened by:	**confounds, small samples, or measurement error** — Confounds, small samples, and error all limit conclusions.
The reason limits matter is that they:	**tell the reader how far to trust the conclusion** — Stated limits calibrate how much weight the conclusion should carry.
Honest scientists report conclusions with:	**appropriate confidence and stated uncertainty** — Match your confidence to the evidence and state the uncertainty.
A responsible conclusion distinguishes:	**what the data shows from what you think it might mean** — Separate the findings from speculation about their meaning.
The right response to an unexpected result is to:	**report it honestly and consider why** — Honest reporting of surprises is essential; investigate the cause.
A single study concluding 'X definitely cures Y' is:	**overclaiming (one study can't prove that)** — One study can't establish a definite cure — that's overclaiming.

16. Design a full investigation + defend it

✂ Question (fold →)	Answer
To reduce the role of chance, your design should have:	**an adequate sample size and replication** — Enough subjects and repetition separate signal from noise.
After the question, you should identify:	**the independent, dependent, and controlled variables** — Naming the variables defines what you'll change, measure, and hold constant.
If your investigation touches a CONTESTED social issue, you should:	**focus on method and evidence, not advocate a position** — Keep to method and evidence; don't turn a study into advocacy (anti-evangelism).
A good investigation report lets another scientist:	**reproduce the study from your methods** — Clear methods enable reproduction — a hallmark of good science.
A pilot study before the full investigation helps you:	**catch design flaws early** — A small pilot surfaces flaws before the full run.
A well-designed study anticipates critique by:	**building in controls, randomization, and adequate sampling from the start** — Strong design pre-empts critique with good controls, randomization, and sampling.
The strongest defense of a causal conclusion is:	**a well-controlled, randomized, replicated design** — Control + randomization + replication is the strongest causal case.
To make your study credible to others, you should:	**pre-register your plan and share your methods** — Pre-registration and open methods build credibility and reproducibility.
A complete capstone design includes question, variables, controls, randomization, measurement, analysis plan, and:	**stated limitations** — A full design also honestly states its limitations.
Your DATA-ANALYSIS plan should be decided:	**before collecting the data** — Deciding the analysis first prevents result-driven (biased) choices.
When presenting results, you should report:	**the effect size and the uncertainty, not just a headline** — Report effect size and uncertainty for an honest picture.
To ensure fairness, treatment and control groups should be:	**treated identically except for the independent variable** — Only the IV should differ between groups.
If a peer critiques your sample as unrepresentative, the honest response is to:	**acknowledge the limit and qualify your generalization** — Acknowledge valid critiques and qualify claims accordingly.
Defending your design means being ready to answer:	**'How do you know a confound didn't cause this?'** — You must show your design rules out confounds and alternatives.
To avoid expectation bias, you might use:	**blinding (and a placebo where relevant)** — Blinding and placebos remove expectation/observer bias.
The first step in designing an investigation is to:	**state a clear, testable question** — A testable question anchors the whole design.
	protecting participants and avoiding harm — Ethical investigations

Ethical design requires:	protect participants and avoid harm.
Your conclusion should be:	**limited to what your data and design support** — Keep the conclusion within the evidence and design's reach.
The synthesis of InquiryForge is:	**good science is a disciplined, honest process of asking, testing, and defending claims with evidence** — Disciplined, honest inquiry — question, test, conclude within the evidence, and defend it — is the core.
The very first thing to pin down when designing an investigation is:	**a clear, testable question** — Everything in the design flows from a clear testable question.
To make your outcome measurable, you must:	**operationalize the dependent variable** — Operationalizing the DV makes the outcome concretely measurable.
The single most important habit this app teaches is:	**following the evidence honestly, wherever it leads** — Honest, evidence-following inquiry is the central habit.
Defending your conclusion means showing it:	**follows from the evidence and survives alternative explanations** — A defensible conclusion follows from evidence and rules out rivals.
To defend against confounds, you should:	**control known variables and randomize to balance unknown ones** — Controlling knowns and randomizing balances confounds.
To support a CAUSAL claim, your design should include:	**random assignment to treatment and control groups** — Random assignment with a control group enables causal inference.

Keep going

You practiced **400 cards** from InquiryForge. Come back to the ones you missed — spaced-out practice makes them stick.

Spark & Anvil is a 501(c)(3) public charity. Every app, story, and book is free, forever — no ads, no in-app purchases, no tracking.

Keep exploring online

Play it now	More free books
	
`https://spark-and-anvil.org/play/inquiryforge`	`https://spark-and-anvil.org/books`

Scan a code with any phone camera, or type the address. Every app, story, and book is free, forever — no account, no tracking.