The Statistics Project Pipeline

A guide to progress, iteration, and uncertainty

A statistics project is **iterative**, not linear. You will revisit earlier stages many times. That is normal—and expected.

---

1. Context & Motivation

What this looks like

- Learning about the real-world problem
- Understanding the domain (science, social science, economics, etc.)
- Asking “What decisions or questions does this data inform?”

Real progress

- A clear description of the context
- Identification of stakeholders or goals

Common trap

- Jumping to methods before understanding the problem

---

2. Research Question Formation

What this looks like

- Translating a real-world problem into statistical questions
- Defining outcomes, predictors, comparisons, or associations

Real progress

- A specific, answerable research question
- Clear population of interest

Common trap

- Questions that are descriptive when inference is intended (or vice versa)

---

3. Data Acquisition & Understanding

What this looks like

- Finding, collecting, or cleaning data
- Understanding how the data were generated

Real progress

- Knowing what each variable represents
- Identifying missingness, bias, or measurement issues

Common trap

- Treating the dataset as “given” rather than constructed

---

4. Exploratory Data Analysis (EDA)

What this looks like

- Visualizations
- Summary statistics
- Checking distributions and relationships

Real progress

- Meaningful plots with interpretation
- New questions emerging from the data

Common trap

- Producing many plots without insight or narrative

---

5. Modeling & Method Selection

What this looks like

- Choosing appropriate statistical models or tests
- Justifying assumptions

Real progress

- Clear rationale for chosen methods
- Awareness of alternative approaches

Common trap

- Picking the “most advanced” method rather than the right one

---

6. Fitting, Checking & Iteration

What this looks like

- Model fitting
- Diagnostics, residuals, validation
- Revising models or variables

Real progress

- Understanding where the model works and fails
- Improving interpretability or fit

Common trap

- Treating model output as truth instead of evidence

---

7. Interpretation & Limitations

What this looks like

- Translating results into plain language
- Acknowledging uncertainty, bias, and assumptions

Real progress

- Clear answers to the research question
- Honest discussion of limitations

Common trap

- Overstating conclusions or hiding uncertainty

---

8. Writing & Communication

What this looks like

- Writing for a non-statistical audience
- Clear figures, tables, and explanations

Real progress

- Results that can be understood without code
- Logical flow from question → method → conclusion

Common trap

- Letting software output replace explanation

---

9. Presentation & Defense

What this looks like

- Explaining choices and tradeoffs
- Responding to questions about data and assumptions

Real progress

- You can explain *why* you did what you did
- You can say what you would do next with more time or data

Common trap

- Believing a “perfect” model exists

Important Reminders

- Data analysis **creates questions** as often as it answers them
- Model revision is a feature, not a failure
- Transparency beats sophistication

Right Now

I am currently in Stage: _____

My next concrete action is:

One key uncertainty I need to address is:
