Scientific Computing in Python Lecture 2

Functions, control flow, and threshold statistics

Lecture notes

A threshold calculation is a useful setting for functions and loops because its mathematical definition is short and its exceptional cases are visible. We shall distinguish a known measurement from a missing value, count the observations satisfying a stated inequality, and identify the first recorded qualifying time. The inequality and the denominator are parts of the statistic.

1 A three-valued classification rule

Slides for this section

Let x be a reading and let θ be a fixed threshold. A missing reading has no numerical order relative to θ and must be handled before comparison. For a known reading the two cases $x \le \theta$ and $x > \theta$ are disjoint and exhaustive. Equality belongs to the first case. The function implements precisely this partition.

$$c(x;\theta) = \begin{cases} \text{missing}, & x \text{ absent}, \\ \text{at or below}, & x \le \theta, \\ \text{above}, & x > \theta. \end{cases}$$

At $\theta = 40$, the values 40 and 40.1 belong to different classes.

A function implementing the partition

```
def classify(value, target):
    if value is None:
        return "missing"
    if value <= target:
        return "at_or_below"
    return "above"
```

2 Arguments, returns, and local names

Slides for this section

A function receives values through its parameters and produces a value with return. A return statement ends that invocation immediately. Local variables belong to the current call. Printing a result displays text but does not replace returning a result for use in another calculation. A function that depends only on explicit arguments is easier to reason about than one whose threshold is stored in a mutable global variable.

```
category = classify(35, 40)
print(category)
```

The returned string is a value. The function can be called again with a different threshold without changing its definition.

Boundary cases

Reading	Threshold	Category
`None`	40	missing
40	40	at or below
40.1	40	above

Testing the boundary establishes whether the implemented inequality agrees with the definition.

3 Indicators and accumulated counts

Slides for this section

For known readings define $I_i(\theta) = 1$ when $x_i \le \theta$ and zero otherwise. The qualifying count is the sum of these indicators. A loop calculates that sum one contribution at a time. After processing the first k observations, the three counters must sum to k. This is a loop invariant: initialization makes it true, and each iteration preserves it by incrementing exactly one counter.

$$N_{\le}(\theta) = \sum_{i:\, x_i \text{ known}} \mathbf{1}_{\{x_i \le \theta\}}.$$

After k observations,

$$N_{\le}^{(k)} + N_{>}^{(k)} + N_{\text{missing}}^{(k)} = k.$$

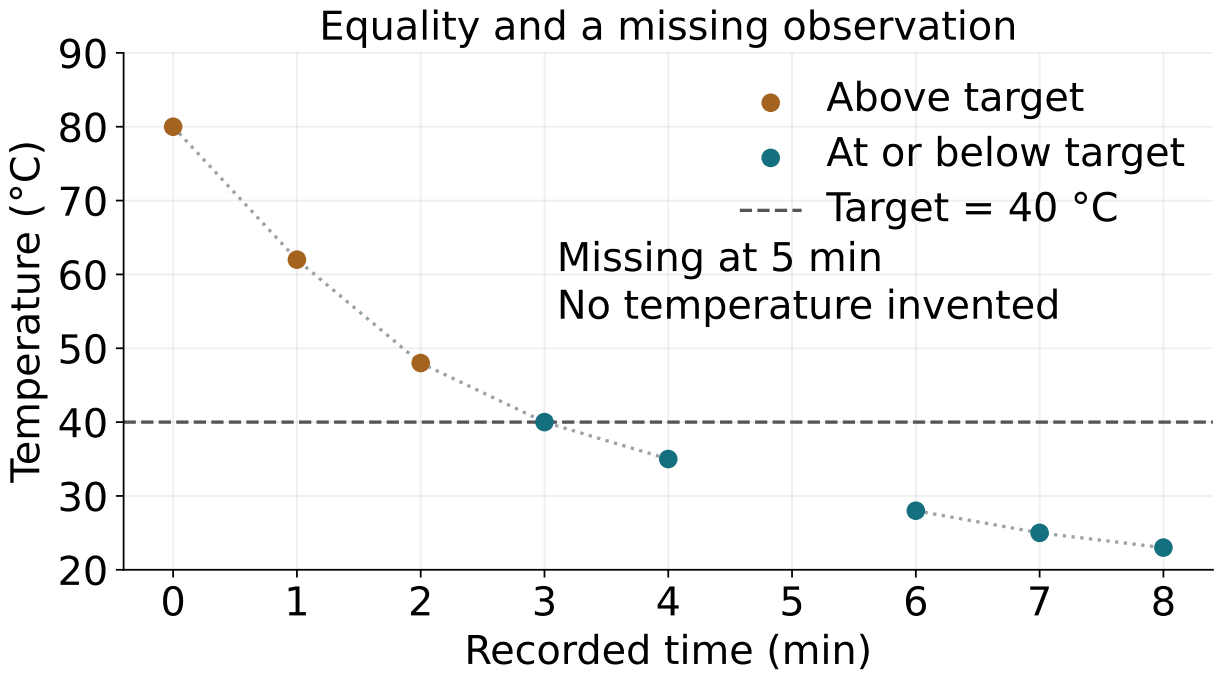
The nine-reading example

$$80, 62, 48, 40, 35, \text{missing}, 28, 25, 23.$$

At threshold 40:

$$N_{\le} = 5, \quad N_{>} = 3, \quad N_{\text{missing}} = 1.$$

The counter identity gives $5 + 3 + 1 = 9$.



The equality case at minute 3 qualifies; the missing reading at minute 5 is left absent. Dotted connectors are visual aids only.

4 The denominator of a proportion

Slides for this section

A proportion must identify its reference group. Among known readings the qualifying fraction is $N_{\leq}/N_{\text{known}}$, where $N_{\text{known}} = N_{\leq} + N_{>}$. Counting a missing reading as a failure silently changes the reference group. If every value is missing the denominator is zero and the proportion is undefined. The function should represent that case explicitly.

$$f(\theta) = \frac{N_{\leq}(\theta)}{N_{\text{known}}}, \qquad N_{\text{known}} > 0.$$

For the supplied record,

$$f(40) = \frac{5}{8} = 0.625.$$

Changing the threshold

At threshold 35, only $35, 28, 25, 23$ qualify:

$$f(35) = \frac{4}{8} = 0.5.$$

The known-reading denominator remains 8 because the data and missingness did not change.

5 The first recorded qualifying time

Slides for this section

The first recorded qualifying time is the minimum of a subset of observation times. An empty subset has no minimum, so we use None to represent the absence of a qualifying reading. Time zero is a valid result and must not be confused with absence. Once the first time has been assigned, subsequent qualifying observations must leave it unchanged.

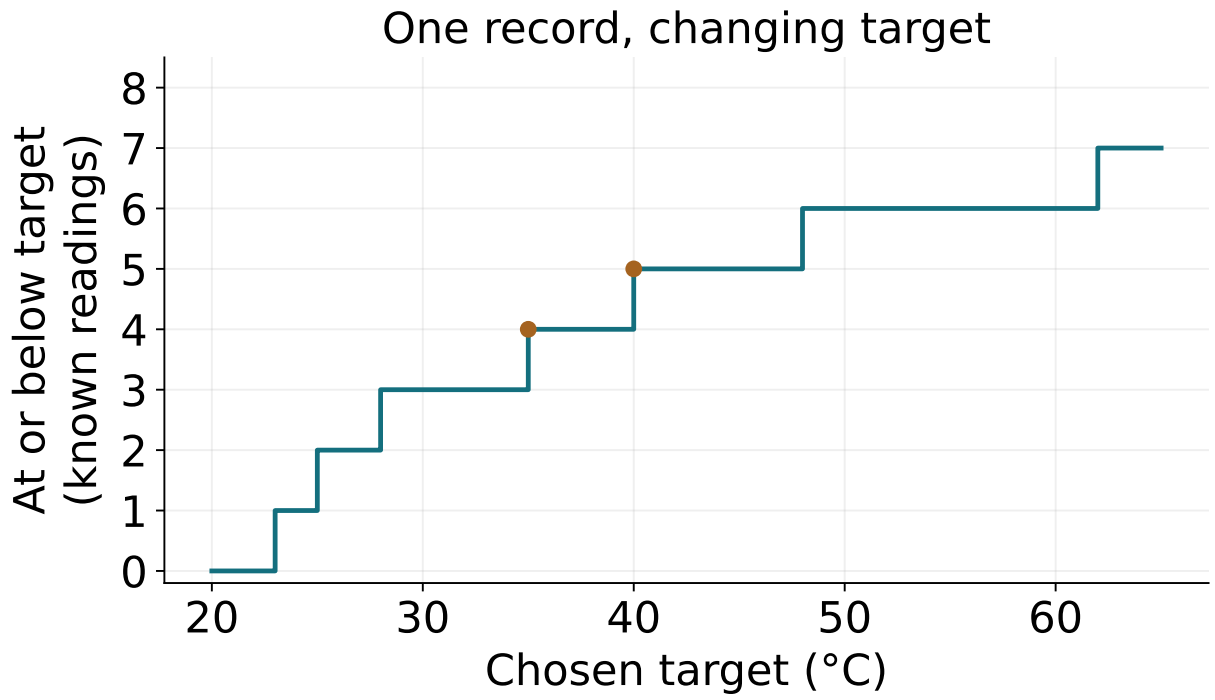
$$t_{\text{first}} = \min\{t_i : x_i \text{ known and } x_i \le \theta\}.$$

For observations at integer minutes and $\theta = 40$, $t_{\text{first}} = 3$ min.

Recorded time and continuous crossing

The record contains $T(2) = 48$ and $T(3) = 40$.

The first qualifying observation occurs at 3 min. An exact continuous crossing requires a model between observations. With oscillation, even earlier unobserved crossings are possible.



The at-or-below count is nondecreasing as the target increases; data and missingness remain fixed.

6 Monotonicity of threshold counts

Slides for this section

If $\theta_1 \le \theta_2$, every observation qualifying at θ_1 also qualifies at θ_2. Thus the qualifying count is nondecreasing in the threshold. For fixed data it is a step function, with jumps at observed values. This

mathematical property provides a test across many thresholds without specifying every answer separately.

$$\{i : x_i \le \theta_1\} \subseteq \{i : x_i \le \theta_2\} \quad (\theta_1 \le \theta_2).$$

Therefore $N_{\le}(\theta_1) \le N_{\le}(\theta_2)$.

A loop with paired observations

```
for time, value in zip(times, readings):
    category = classify(value, target)
    counts[category] += 1
    if category == "at_or_below" and first is None:
        first = time
```

Equal list lengths must be checked: `zip` stops at the shorter input.

Exercises

H2-1. Trace the decision boundary

Use target 37 degrees Celsius and readings [None, 37, 37.1, 36.9]. New practice; 10 feedback points.
a Give the category of each reading in order for an at-or-below rule.

b State the known-reading denominator and qualifying fraction.

c How do the count and fraction change if the rule is strictly below?

H2-2. Keep the first qualifying time

Times are [0, 2, 4, 6] minutes and temperatures are [35, 44, None, 34] degrees Celsius. Target: 35 degrees. New practice; 10 feedback points.
a Trace the qualifying count and stored first time after each observation.

b Explain the bug in replacing a test of first_time is None with not first_time.

c What can this record establish about an exact continuous crossing time?

H2-3. Write and check a reusable rule

Work with small supplied numeric inputs and None. New practice; 10 feedback points.
a Give the complete ordered decision logic for a function with a reading and a target as inputs. It must return one of missing, at_or_below or above.

b Give three test calls that distinguish boundary, above-boundary and missing behavior, with expected results.

c Explain why printing the label without returning it is insufficient for a later counting function.

H2-4. Interpret missingness and a changing target

A new record contains [12, None, 18, 21, None] degrees Celsius. New practice; 10 feedback points.
a For target 18 degrees, give all category counts and the fraction among known readings.

b Prove in words that raising the target cannot decrease the qualifying count when data are fixed.

c If every reading were None, what should a report say about the qualifying fraction, and why?

References

Original Python Lecture 0. PDF pages 44–45 and 47–49.
https://zanahmad.com/files/scientific-python/Lecture_0__Intro_to_Python.pdf

Original Python Lecture 1. PDF pages 4–9.
https://zanahmad.com/files/scientific-python/Lecture_1__Intro_to_Python.pdf

Python source materials: Phillip Kerger (2023). Course adaptation and additional exposition: Zan Ahmad.