DS 1000B Section 003
Week 2: Errata, remarks & FAQ

Winter 2026

Remarks

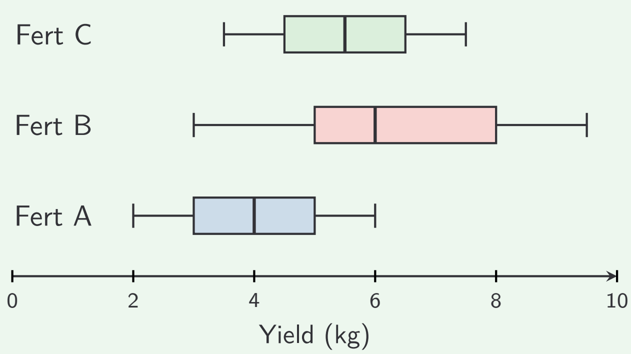
Boxplots interpretation

When interpreting boxplots, you do not have to consider shape (e.g., skewness) as this can be misleading. Instead, focus on median (the center), IQR (the spread) and outliers.

I realize that some of the practice problems in the notes were more challenging than intended. The following are solutions to some of these problems for your reference.

Exercise 2.55

An agricultural researcher tested three different fertilizers (A, B, and C) on tomato plants and recorded the yield (in kg) per plant. The results are summarized in the boxplots below.



(a) Which fertilizer produced the highest typical yield?
(b) Which fertilizer produced the most consistent results (lowest variability)?
(c) Describe the shape of the distribution for Fertilizer B.
(d) If you wanted to guarantee a yield of at least 4 kg per plant, which fertilizer would you choose and why?

(a) Fertilizer B. It has the highest median yield (approx. 6 kg) compared to Fertilizer C (5.5 kg) and Fertilizer A (4 kg).

(b) Fertilizers A and C. Both have the smallest Interquartile Range (IQR ≈ 2 kg) and overall range (4 kg), indicating the lowest variability. Fertilizer B is the most variable.

(c) Skewed Right (Positively Skewed). (See remark above)

(d) Fertilizer B. While Fertilizer C is more consistent, Fertilizer B offers the highest probability of exceeding 4 kg. Since $Q1 = 5$ kg for Fertilizer B, approximately 75% of the plants produced at least 5 kg. In contrast, Fertilizer A has a median of 4 kg (meaning 50% of plants produced less than 4 kg).

Note: the following problem is too advanced for the current course scope, but I have included it for those that were interested in a challenge.

Exercise 2.56

A dataset has a median of 60. You are told that the value **112** is the *smallest possible integer* that classifies as a high outlier according to the $1.5\times$IQR rule (i.e., $x >$ Upper Fence). Using this information, determine the theoretical maximum possible value for the third quartile (Q_3).

To maximize Q_3, we must maximize the Upper Fence and maximize Q_1.

1. Outlier Condition: A value x is a high outlier if $x >$ Upper Fence, where Upper Fence $= Q_3 + 1.5(IQR)$.
2. Fence Constraint: We are told 112 is the *smallest* integer outlier. This implies the fence must be strictly less than 112 (so 112 is an outlier) but greater than or equal to 111 (so 111 is not an outlier). To maximize Q_3, we use the upper bound:

$$\text{Upper Fence} < 112$$

3. Q_1 Constraint: By definition, $Q_1 \leq$ Median. Given the Median is 60, the maximum possible value for Q_1 is 60.
4. Putting these together, we find

$$Q_3 + 1.5(Q_3 - Q_1) < 112$$

$$2.5Q_3 - 1.5Q_1 < 112$$

As $Q1 \leq 60$, we find

$$\begin{aligned} 2.5Q_3 - 1.5Q_1 < 2.5Q_3 - 1.5(60) &< 112 \\ 2.5Q_3 - 90 &< 112 \\ 2.5Q_3 &< 202 \\ Q_3 &< 80.8 \end{aligned}$$

Answer: The theoretical maximum value for Q_3 is just below 80.8.

Frequently Asked Questions