Data Set B – Diets

Analysis of Diet A vs Diet B

Diet A	n	50
	Mean	5,341
	SD	2,536
	Median	5,642
	Q1	3,748
	Q3	7,033
	IQR	3,285
Diet B	n	50
	Mean	3,710
	SD	2,769
	Median	3,745
	Q1	1,953
	Q3	5,404
	IQR	3,451

Sample Size (n = 50) is the same for both diets, allowing a fair comparison.

Central Tendency:

- Diet A has a higher mean weight loss (5.341 kg) compared to Diet B (3.710 kg), suggesting that participants on Diet A lost more weight on average.

- The median weight loss for Diet A (5.642 kg) is also greater than for Diet B (3.745 kg), reinforcing that Diet A performed better for most participants.

Spread and Variability:

- Standard deviation (SD) is slightly lower for Diet A (2.536) than for Diet B (2.769), indicating that weight loss results were more consistent for Diet A.

- The interquartile range (IQR) is also smaller for Diet A (3.285) than for Diet B (3.451), which suggests that the middle 50% of results were less spread out in Diet A.

- Q1 and Q3 values confirm this pattern: both the lower and upper quartiles are higher for Diet A, showing that even participants with lower results still did relatively better on this diet.

Conclusion:

Based on the statistical measures alone, Diet A appears more effective and consistent in producing weight loss than Diet B. It not only shows higher average and median losses but also less variability among participants.

Diet A			UCB	Frequency		Class Mark	Relative Frequency
	n	50	0	1		-1	2,00%
	Mean	5,341	2	3		1	6,00%
	SD	2,536	4	10		3	20,00%
			6	15		5	30,00%
	Min	-1,715	8	15		7	30,00%
	Max	10,062	10	5		9	10,00%
	Range	11,777	12	1		11	2,00%
			Total	50			
						Total RF	100,00%

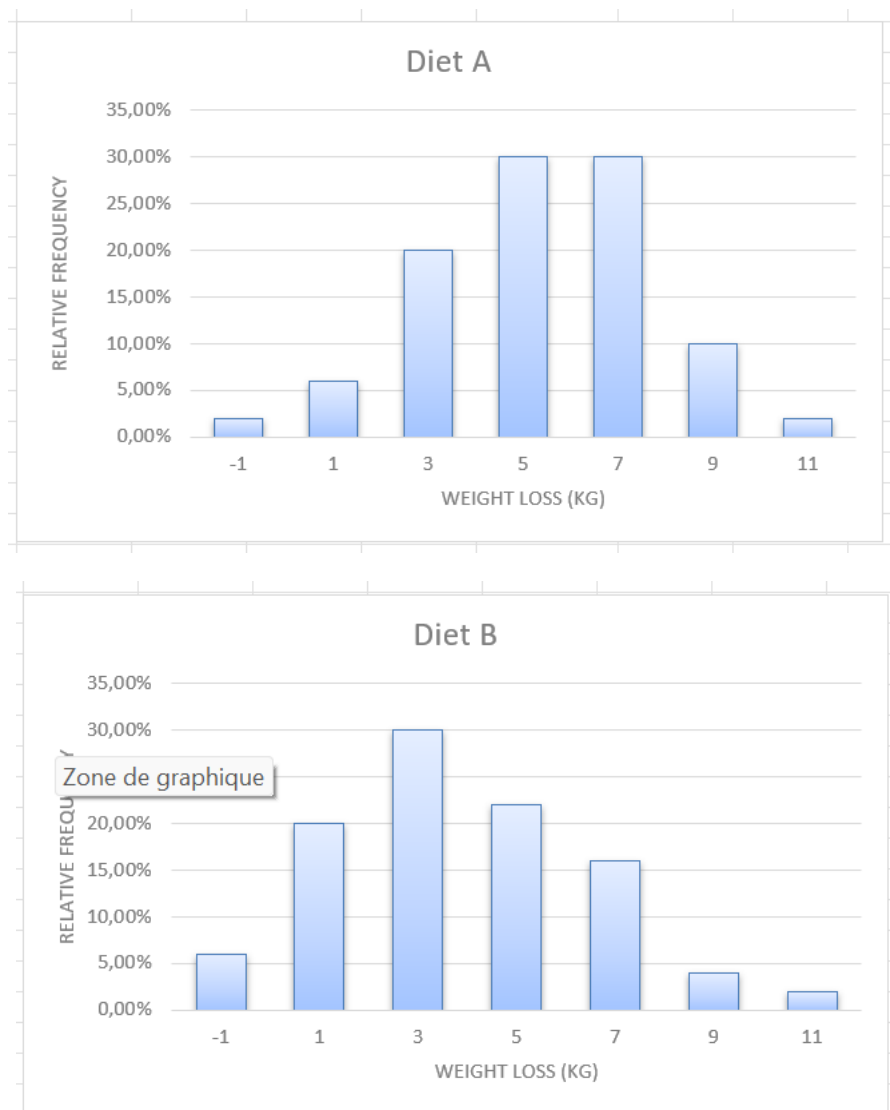| Diet B			UCB	Frequency		Class Mark	Relative Frequency
	n	50	0	3		-1	6,00%
	Mean	3,710	2	10		1	20,00%
	SD	2,769	4	15		3	30,00%
	Min	-4,148	6	11		5	22,00%
	Max	10,539	8	8		7	16,00%
	Range	14,687	10	2		9	4,00%
			12	1		11	2,00%
			Total	50			
						Total RF	100,00%

Distribution Shape:

- **Diet A:**

  - The most common weight loss values lie in the 5–7 kg range (Class Marks 5 and 7), each with 30% of the data.

  - This suggests a strong central tendency and a symmetrical distribution centered around moderate to high weight loss.

  - Very few participants are at the extremes (e.g., only 2% lost ~11 kg, and 1 person had weight gain or minimal loss).

- **Diet B:**

  - The most frequent weight loss occurred at 3 kg (30%), followed by 5 kg (22%) and 1 kg (20%).

  - Diet B's distribution is skewed to the left (negative skew) — meaning more participants had lower weight loss.

  - 6% of participants were in the weight gain or no loss range (~-1 kg) — higher than Diet A.

Interpretation:

- Diet A shows a stronger and more consistent effect, with a significant majority (60%) losing between 5 and 7 kg.

- Diet B, while showing some positive results, has more spread and more participants experiencing minimal benefit.

- This supports the idea that Diet A is both more effective and more reliable, even without visualizing the data.

Interpretation of Visual Data

Diet A histogram: Relative Frequency (%) vs Weight Loss (kg). Bars at -1: 2%, 1: 6%, 3: 20%, 5: 30%, 7: 30%, 9: 10%, 11: 2%.

Diet B histogram: Relative Frequency (%) vs Weight Loss (kg). Bars at -1: 6%, 1: 20%, 3: 30%, 5: 22%, 7: 16%, 9: 4%, 11: 2%. Tooltip labeled "Zone de graphique".

Diet A:

- The distribution is fairly symmetrical, peaking between 5–7 kg.

- Most participants experienced moderate to high weight loss.

- Very few participants had minimal or negative weight changes.

- This supports the idea that Diet A is both effective and consistent.

Diet B:

- The distribution is skewed slightly left, with more data concentrated between 3–5 kg.

- Although some participants experienced significant weight loss, others had minimal or even negative changes (weight gain).

- The spread is wider, reflecting greater variability in outcomes.

- Diet B appears less reliable overall.

Conclusion:

The visual evidence confirms what the descriptive statistics suggested: Diet A delivers better and more consistent weight loss results compared to Diet B. The concentration of higher relative frequencies in the mid-to-upper weight loss ranges for Diet A reinforces its effectiveness.

While Diet A led to greater average weight loss, its slightly lower variability may make it more suitable for programmes that emphasise consistent outcomes. Diet B, being more variable, might be preferable where flexibility or less intensive approaches are desired. It may also be appropriate in contexts where personalisation is prioritised.

Data Set D – Brandprefs

Frequencies

	Area 1	Area 2
A	11	19
B	17	30
Other	42	41
Total	70	90

Percentages

	Area 1	Area 2
A	15,7	21,1
B	24,3	33,3
Other	60,0	45,6
Total	100,0	100,0

The frequency and percentage data reveal notable differences in brand preferences between the two demographic areas.
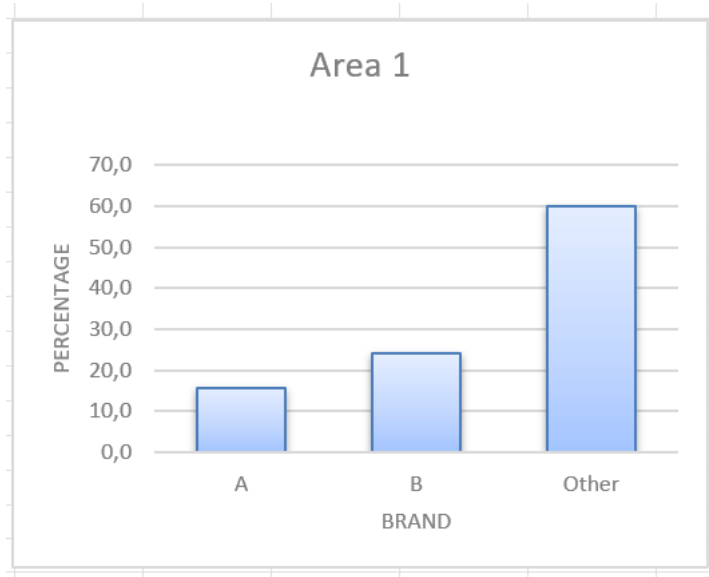
In Area 1:

- Brand A was chosen by 11 out of 70 participants, representing 15.7%.

- Brand B was preferred by 17 participants (24.3%).

- A significant 60.0% (42 participants) selected Other brands, indicating that the majority do not favor either of the manufacturer's key brands in this area.
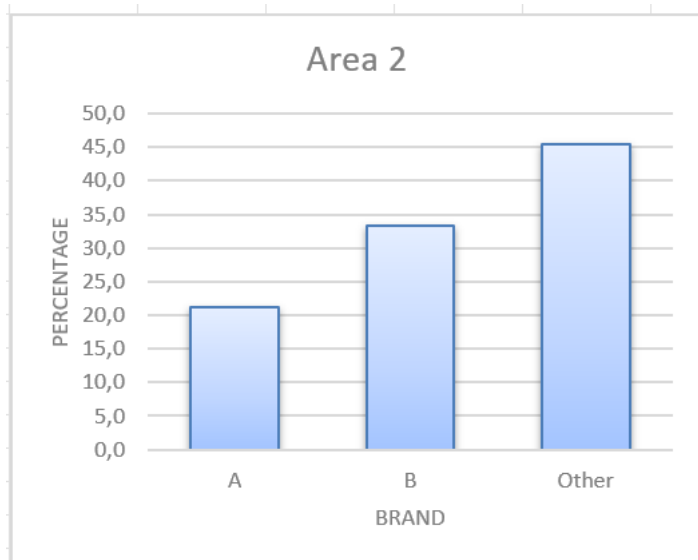
In Area 2:

- Brand A was chosen by 19 out of 90 participants (21.1%).

- Brand B was selected by 30 participants (33.3%), the highest proportion for any brand in either area.

- 41 participants (45.6%) chose Other brands, a lower proportion than in Area 1.

This comparison shows that Area 2 has a stronger preference for the manufacturer's brands, particularly Brand B, while Area 1 shows a clear tendency toward alternative brands. These differences suggest that brand recognition or loyalty varies significantly by demographic region, with greater marketing challenges or competitor influence in Area 1.

Visual interpretation

The bar charts for Area 1 and Area 2 illustrate the percentage distribution of brand preferences across three categories: Brand A, Brand B, and Other.

In Area 1, the chart clearly shows:

- A dominant bar for "Other", reaching 60%, indicating a strong preference for brands outside of A and B.

- Brand B appears as the second-highest, around 24%, with Brand A having the lowest bar at approximately 16%.

- The visual gap between "Other" and the named brands is substantial, highlighting the weak presence of the manufacturer's brands in this area.

In Area 2, the distribution is more balanced:

- "Other" is still the highest, but now only around 46%, significantly lower than in Area 1.

- Brand B shows a strong presence, with a bar close to 33%, making it the most preferred named brand.

- Brand A also shows a modest increase, rising to about 21%.

These visual differences reinforce the conclusion that:

- Area 2 is more favorable toward the manufacturer's brands, particularly Brand B.

- Area 1 presents a challenge, as consumers there are clearly more inclined toward alternative brands, visually emphasized by the height of the "Other" bar.

The higher preference for branded products in Area 2 (especially Brand B at 33.3%) suggests this demographic may respond well to targeted advertising or loyalty

programmes. In contrast, the dominance of 'Other' in Area 1 (60%) hints at either brand disinterest or sensitivity to price. This may indicate a preference for local or budget-conscious options.

Data Set C – Superplus

Descriptive Interpretation

This data summarizes the annual income (in £) of 60 male and 60 female cardholders.

| Men	n	60
	Mean	52,9
	SD	15,269
	Median	52,1
	Q1	40,75
	Q3	59,2
	IQR	18,45
Women	n	60
	Mean	44,2
	SD	13,790
	Median	38,2
	Q1	33,325
	Q3	52,725
	IQR	19,4

Central Tendency:

- Men have a higher mean income (£52.9k) compared to women (£44.2k), a difference of £8.7k.

- The median income for men (£52.1k) is also notably higher than for women (£38.2k), indicating that even the typical male earner receives more.

Spread and Variability:

- Standard deviation is greater for men (15.269), suggesting more variability in male incomes.

- Women show slightly wider IQR (19.4 vs 18.45), which means the middle 50% of their incomes are more spread out.

- Quartile values reinforce this: women's lower quartile (Q1 = £33.3k) is below that of men (Q1 = £40.8k), and their upper quartile (Q3 = £52.7k) is also lower than that of men (Q3 = £59.2k).

Visual Interpretation
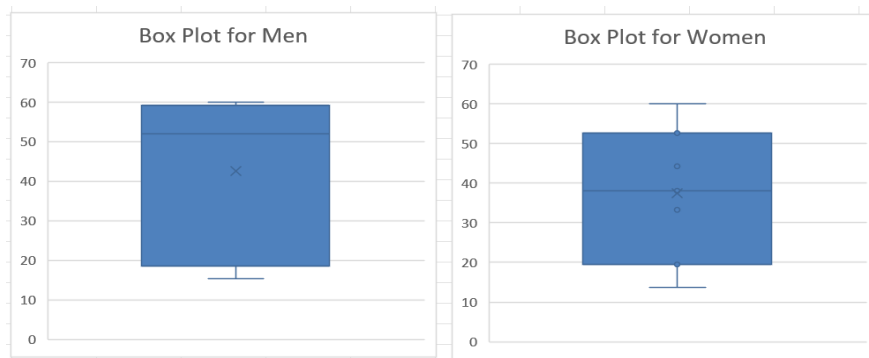
The side-by-side box plots offer a clear visual comparison between the two groups:

Men:

- The box is positioned higher on the income scale, consistent with the higher median and quartile values.

- The median line is close to the top of the box, suggesting a slight negative skew.

- Fewer visible outliers indicate more consistency in extreme income values.

Women:

- The box plot is lower overall, reflecting lower income levels.

- A more evenly spaced distribution between quartiles is shown, but with several outliers, indicating income extremes in the female group.

- The spread is wide, with the lower whisker extending further down than for men.

Conclusion:

The box plots reinforce the descriptive findings: men generally earn more, and their incomes are more clustered around higher values. Women's incomes are more varied and show a greater presence of lower and outlier values, hinting at more diverse financial situations within that group. These findings may reflect broader structural or societal factors, such as differences in job roles, working hours, or industry representation between genders. While the data alone do not explain causation, they suggest that income disparity could stem from more than just individual performance or qualifications.

Data Set G – Filtration

This dataset compares impurity levels (in parts per 1000) remaining after filtration using two different agents.

|	Agent1	Agent2
n	12	12
Mean	8,25	8,68
SD	1,03	1,04
Median	8,4	9,0
Q1	7,4	8,1
Q3	9,25	9,45
IQR	1,85	1,38

Central Tendency:

- Agent 1 has a lower mean impurity level (8.25) than Agent 2 (8.68), suggesting it may be slightly more effective on average.

- The medians reinforce this: Agent 1 = 8.4 vs Agent 2 = 9.0, again in favor of Agent 1.

Spread and Variability:

- Both agents have similar standard deviations (Agent 1 = 1.03, Agent 2 = 1.04), showing comparable variability.

- The IQR is larger for Agent 1 (1.85 vs 1.38), indicating that its middle 50% of values are more spread out.

- Quartile comparisons:

    o Q1 for Agent 1 is 7.4 vs 8.1 for Agent 2

    o Q3 for Agent 1 is 9.25 vs 9.45 for Agent 2

This means that the bulk of values for Agent 1 are slightly lower overall, but also more variable within the central range.

Visual Interpretation
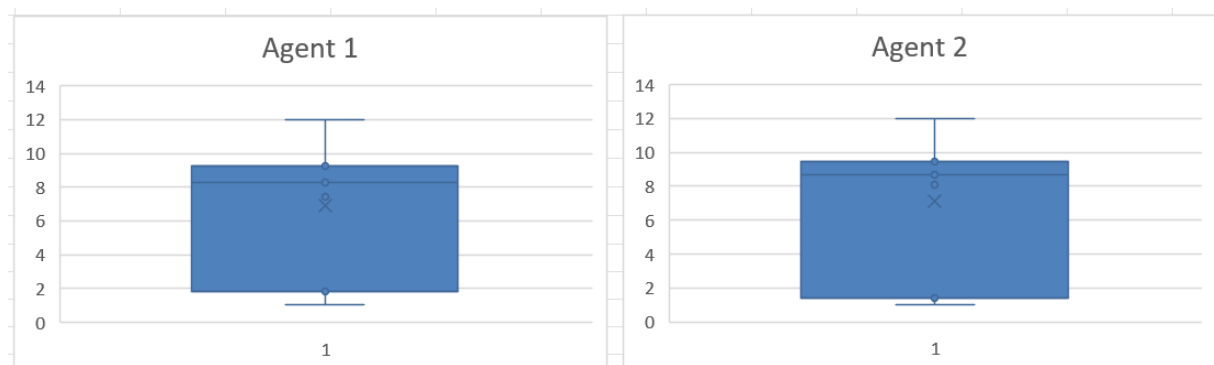
- ◆ **Agent 1:**

- The median line is closer to the top of the box, indicating a slightly left-skewed (negatively skewed) distribution.

- There are several outliers, including a low outlier near 0, which suggests a possible occasional filtration anomaly.

- The overall impurity range is wider compared to Agent 2.

Agent 2:

- The box is more symmetrical, with the median near the center.

- Outliers are present as well, but its distribution appears slightly more stable.

- Despite the slightly higher central values, its IQR is narrower, indicating more consistency.

Conclusion:

While Agent 1 shows slightly better performance on average (lower mean and median impurity), its results are more variable, with some extreme low values. Agent 2, on the other hand, offers more consistent outcomes, though with slightly higher average impurity levels. The best choice may depend on whether lower impurity or process consistency is the priority. If this were a pharmaceutical or food safety context, achieving the lowest possible impurity level might be the priority, making Agent 1 more attractive despite its variability. However, in industries where predictable performance is essential, the consistency offered by Agent 2 might be preferred.