Curtin University, Semester 2, 2026

ECOM5005

Tutorial 3 Solutions

Numerical Descriptive Measures - classroom-ready worked solutions

Learning Outcomes

1. Describe, calculate and interpret numerical descriptive measures of central tendency, variation, and shape for numerical data.
2. Describe, calculate and interpret descriptive summary measures for a population.
3. Construct and interpret a box-and-whisker plot.
4. Calculate and interpret the covariance and the coefficient of correlation for bivariate data.

PART 1 - Detailed worked solutions

Question 1.

Question

Given the grouped frequency distribution of defective tomatoes in 50 shipments, find the approximate mean number of defective tomatoes.

Detailed Solution

Because the raw observations are not available, we approximate every value in a class by the midpoint of that class. For example, the midpoint of 0 to less than 5 is (0 + 5)/2 = 2.5.

Multiply each class midpoint by its frequency. These products estimate the total number of defective tomatoes represented by each class. Then add the products and divide by the total frequency.

Defective items	Frequency f	Midpoint m	f × m
0 to <5	35	2.5	87.5
5 to <10	7	7.5	52.5
10 to <15	5	12.5	62.5
15 to <20	2	17.5	35.0
20 to <25	1	22.5	22.5
25 to <30	0	27.5	0.0
Total	50		260.0

Approximate mean = $\Sigma(f \times m) / \Sigma f = 260 / 50 = 5.2$ defective tomatoes.

Interpretation

The estimated average is approximately 5.2 defective tomatoes per shipment. It is an approximation because all observations within each interval are represented by the class midpoint.

Question 2.

Question

Compare mean and median in two settings: distances of bright celestial objects and annual income.

(a) Celestial-body distances

Rank	Object	Distance (light years)
1	Moon	0.000000038
2	Venus	0.000048
3	Jupiter	0.000067
4	Mars	0.000076
5	Mercury	0.000095
6	Saturn	0.0014
7	Syrius	8.6
8	Canopus	310

Detailed Solution

Mean: add all eight distances and divide by 8. The total is approximately 318.601686038 light years, so the mean is 318.601686038 / 8 = 39.825210755 light years.

Median: there are 8 observations, so take the average of the 4th and 5th ranked values: (0.000076 + 0.000095)/2 = 0.0000855 light years.

The median is the more informative measure of a typical distance because the distribution is extremely right-skewed. Canopus (310 light years) and, to a lesser extent, Syrius (8.6 light years), pull the arithmetic mean far above most observations.

Answer Summary

Mean ≈ 39.8252 light years; Median = 0.0000855 light years; use the median to communicate the typical distance.

(b) Income

Detailed Solution

From the mean alone, the exact median cannot be determined. However, income distributions are commonly right-skewed because a small number of very high incomes can pull the mean upward.

Therefore, a reasonable expectation is that the sample median would be less than \$50,000. This is a likely pattern, not a mathematical certainty without the full data.

Question 3.

Question

Analyse the 15 bank waiting times using measures of centre, spread, Z-scores, outliers and shape, and evaluate the manager's claim that a customer will almost certainly wait less than five minutes.

Step 1: Order the data

0.38, 2.34, 3.02, 3.20, 3.54, 3.79, 4.21, 4.50, 4.77, 5.12, 5.13, 5.55, 6.10, 6.19, 6.46

Step 2: Measures of centre and quartiles

Worked calculations

Mean = 64.30 / 15 = 4.2867 minutes.

Median position = (15 + 1)/2 = 8, so the median is the 8th observation = 4.50 minutes.

Q1 position = (15 + 1)/4 = 4, so Q1 = 3.20 minutes.

Q3 position = 3(15 + 1)/4 = 12, so Q3 = 5.55 minutes.

Step 3: Measures of variation and shape

Measure	Result
Mean	4.2867
Median	4.5000
Q1	3.2000
Q3	5.5500
Sample variance	2.6830
Sample standard deviation	1.6380
Range	6.0800
Interquartile range	2.3500
Coefficient of variation	38.21%
Skewness	-0.833

The IQR is Q3 - Q1 = 5.55 - 3.20 = 2.35. The coefficient of variation is (1.637985 / 4.286667) × 100 ≈ 38.21%.

Step 4: Z-scores

Waiting time	Z-score
4.21	-0.05
5.55	0.77
3.02	-0.77
5.13	0.51
4.77	0.30
2.34	-1.19
3.54	-0.46
3.20	-0.66
4.50	0.13
6.10	1.11
0.38	-2.39
5.12	0.51
6.46	1.33
6.19	1.16
3.79	-0.30

Outlier conclusion

Lecture 3 treats an observation as an extreme outlier when Z < -3 or Z > +3. The largest absolute Z-score is about 2.39 for the 0.38-minute wait, so there are no extreme Z-score outliers.

The 1.5×IQR fences are -0.325 and 9.075 minutes, which also place all observations inside the fences.

Figure 1: Boxplot of the 15 bank waiting times

Shape and manager claim

The sample skewness is approximately -0.833, so the distribution is moderately left-skewed, largely because the 0.38-minute observation creates a longer lower tail. The fact that the mean (4.287) is below the median (4.50) is consistent with this.

Six of the 15 customers waited more than five minutes: 5.12, 5.13, 5.55, 6.10, 6.19 and 6.46. Thus 40% of the sample exceeded five minutes. The manager is correct that a wait below five minutes is more likely than not, but saying "almost certainly" is not supported by this sample.

Question 4. Investment Performance and Risk - two managed funds

Question

Compare the six-year returns of Fund A and Fund B using arithmetic mean, geometric mean, standard deviation, coefficient of variation and ending wealth.

Year	Fund A	Fund B
1	12%	8%
2	-8%	7%
3	18%	9%
4	5%	6%
5	10%	8%
6	3%	7%

Detailed Solution

Arithmetic mean return: Fund A = (12 - 8 + 18 + 5 + 10 + 3)/6 = 6.67%. Fund B = (8 + 7 + 9 + 6 + 8 + 7)/6 = 7.50%.

Geometric mean return uses compounded growth factors. Fund A: [(1.12)(0.92)(1.18)(1.05)(1.10)(1.03)]^(1/6) - 1 = 6.35%. Fund B: [(1.08)(1.07)(1.09)(1.06)(1.08)(1.07)]^(1/6) - 1 = 7.50%.

Sample standard deviations: Fund A ≈ 8.94 percentage points; Fund B ≈ 1.05 percentage points.

Coefficient of variation: Fund A ≈ 134.05%; Fund B ≈ 13.98%. The much smaller CV for Fund B shows substantially lower variability relative to its average return.

Ending value of $10,000 after six years: Fund A ≈ $14,464.62; Fund B ≈ $15,429.34.

Measure	Fund A	Fund B
Arithmetic mean	6.67%	7.50%
Geometric mean return	6.35%	7.50%
Sample standard deviation	8.94%	1.05%
Coefficient of variation	134.05%	13.98%

Business interpretation

For a risk-conscious investor, Fund B dominates this six-year sample: it has a slightly higher average and compound return, substantially lower volatility, a much lower coefficient of variation, and a higher ending value. The geometric mean is the appropriate long-run growth measure because annual returns compound multiplicatively.

Measure	Fund A	Fund B
Ending value of $10,000	$14,464.62	$15,429.34

Question 5. Digital Marketing Analytics - advertising spend and online sales

Question

Use the eight weekly observations to examine the linear relationship between digital advertising expenditure and online sales.

Week	Ad spend X ($000)	Sales Y ($000)
1	5	45
2	7	47
3	8	58
4	10	49
5	12	63
6	13	55
7	15	72
8	18	68

Detailed Solution

The mean advertising spend is X-bar = 11.000 ($000) and the mean weekly sales value is Y-bar = 57.125 ($000).

Sample covariance = Σ[(Xi - X-bar)(Yi - Y-bar)]/(n - 1) = 36.5714. The covariance is positive, so advertising spend and sales tend to move in the same direction.

The sample standard deviations are SX ≈ 4.3425 and SY ≈ 9.9633. Therefore r = covariance/(SX × SY) = 36.5714/(4.3425 × 9.9633) ≈ 0.8453.

An r of about 0.845 indicates a strong positive linear association in this small sample: weeks with higher digital advertising expenditure tend to have higher online sales.

Correlation is not causation. Sales may also be affected by discounts, holidays, product launches, competitor activity, website traffic quality, stock availability or seasonality. These factors can move together with advertising and sales.

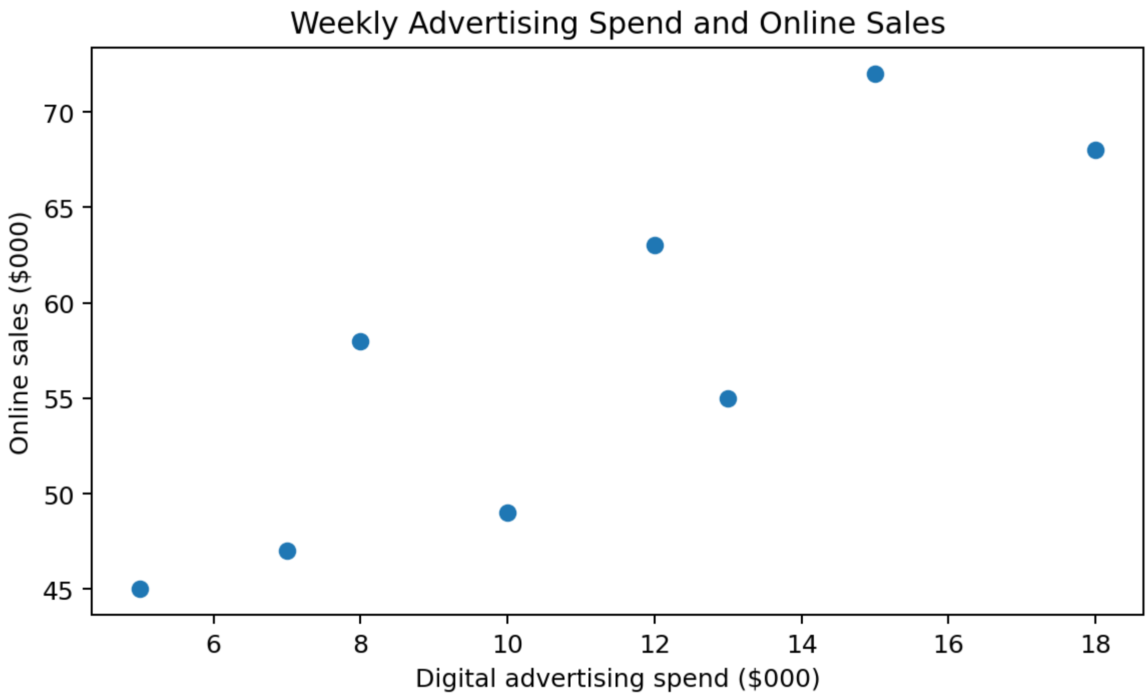

Figure 2 : Scatter diagram for weekly digital advertising spend and online sales

PART 2 - Data exercise: Beer.xlsx

Purpose of this section

This section is written for students with little prior Excel experience. Follow the steps in order. The goal is not only to obtain numbers, but also to create an auditable Excel analysis that another person can reproduce.

Step 1 - Open the file and understand the data

5. Open Beer.xlsx and select the worksheet named Beer.
6. The data occupy A1:B96. Row 1 contains the headings Calories/100ml and Type; rows 2 to 96 contain 95 beer observations.
7. Calories/100ml is numerical and is treated as continuous for analysis. Type is categorical with three categories: Regular, Light and Non-alcoholic.
8. Before calculating anything, save a working copy so that the original data remain unchanged.

Data-check formulas

In an empty cell type =COUNT(A2:A96). The result should be 95.

Type =COUNTBLANK(A2:A96) and =COUNTBLANK(B2:B96). Both should return 0.

Use =COUNTIF(B2:B96,"Regular"), =COUNTIF(B2:B96,"Light") and =COUNTIF(B2:B96,"Non-alcoholic") to verify the three type counts.

Check	Final result
Valid calorie observations	95
Missing Calories/100ml	0
Missing Type	0
Regular	68
Light	18
Non-alcoholic	9

Step 2 - Produce the Excel ToolPak descriptive statistics

9. If Data Analysis is not visible on the Data tab, enable the Analysis ToolPak: File -> Options -> Add-ins -> Manage Excel Add-ins -> Go -> tick Analysis ToolPak -> OK.
10. Choose Data -> Data Analysis -> Descriptive Statistics.
11. Set Input Range to $A$1:$A$96, choose Grouped By: Columns, and tick Labels in first row.
12. Choose an output location on a new worksheet (for example, name it Analysis) and tick Summary statistics.
13. Click OK. Excel will create the standard descriptive-statistics table.

Statistic	Final Excel result
Mean	39.1158
Standard Error	1.1136
Median	42.0000
Mode	43.0000
Standard Deviation	10.8538
Sample Variance	117.8056
Kurtosis	2.5469
Skewness	-0.0053
Range	69.0000
Minimum	14.0000
Maximum	83.0000
Sum	3716.0000
Count	95

Interpret the ToolPak output

The mean calorie content is about 39.12 calories per 100 ml, while the median is 42 and the mode is 43. Overall skewness is almost zero, indicating little net asymmetry in the combined sample.

The positive kurtosis of about 2.55 indicates heavier tails / greater peakedness than the normal benchmark used by Excel, and the value 83 is much higher than most observations.

Step 3 - Add quartiles, IQR and coefficient of variation

Excel formulas that match the lecture convention

First quartile: =QUARTILE.EXC($A$2:$A$96,1) -> 33.

Third quartile: =QUARTILE.EXC($A$2:$A$96,3) -> 45.

Interquartile range: =Q3_cell-Q1_cell -> 12.

Coefficient of variation: =STDEV.S($A$2:$A$96)/AVERAGE($A$2:$A$96)*100 -> 27.7479%. Format the result as a percentage or report 27.75%.

Measure	Result
Q1	33
Q3	45
IQR	12
Coefficient of variation	27.75%

Step 4 - Calculate Z-scores and identify extreme outliers

14. Insert a new column C and label C1 as Z-score.
15. In C2 type =(A2-AVERAGE($A$2:$A$96))/STDEV.S($A$2:$A$96). Press Enter.
16. Copy the formula down to C96 by double-clicking the fill handle at the bottom-right corner of C2.
17. In D1 type Extreme outlier?. In D2 type =IF(ABS(C2)>3,"Extreme outlier","") and fill down to D96.
18. Filter column D for Extreme outlier.

Observation	Calories/100ml	Z-score	Conclusion
Maximum beer	83	4.0432	Extreme high outlier
Minimum beer	14	-2.3140	Not an extreme Z-score outlier

Important distinction

Using the Lecture 3 Z-score rule, 83 is the only extreme outlier because |Z| > 3.

Using the 1.5×IQR boxplot rule for the combined sample gives lower and upper fences of 15 and 63. Under that rule, the three observations equal to 14 and the observation 83 are potential boxplot outliers. Different outlier rules answer slightly different questions.

Step 5 - Build the frequency distribution

19. On the Analysis sheet create three columns headed Class interval, Frequency and Percentage.
20. Use the class intervals 10 to <20, 20 to <30, 30 to <40, 40 to <50, 50 to <60, 60 to <70, 70 to <80 and 80 to <90.
21. For the first class, use =COUNTIFS($A$2:$A$96,">=10",$A$2:$A$96,"<20"). Repeat the same logic for each class.
22. For percentage, divide each class frequency by 95 and format as Percentage with two decimal places. The percentages should sum to 100%.

Class interval	Frequency	Percentage
10 to <20	7	7.37%
20 to <30	11	11.58%
30 to <40	17	17.89%
40 to <50	52	54.74%
50 to <60	6	6.32%
60 to <70	1	1.05%
70 to <80	0	0.00%
80 to <90	1	1.05%
Total	95	100.00%

Step 6 - Create the histogram

23. Select the class labels and frequency values from the frequency table.
24. Choose Insert -> Column or Bar Chart -> 2-D Clustered Column.

25. Right-click a bar -> Format Data Series -> set Gap Width to 0% so the bars touch like a histogram.
26. Add the title Distribution of Calories per 100 ml for 95 Beers. Label the horizontal axis Calories per 100 ml and the vertical axis Frequency.
27. Do not use 3-D effects, unnecessary pictures or decorative chartjunk.

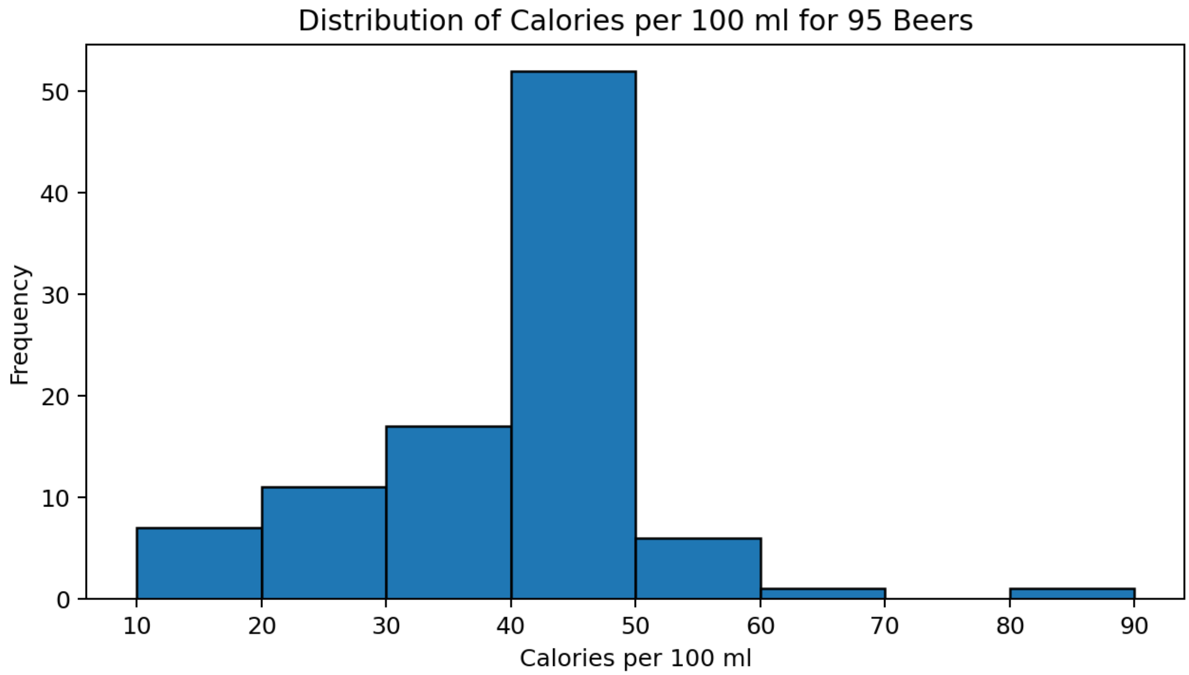

Figure 3: Histogram using 10-calorie class intervals

Histogram interpretation

The dominant interval is 40 to less than 50 calories, containing 52 of 95 beers (54.74%). Most beers are therefore concentrated around the low-to-mid 40s.

There is one observation in the 80 to <90 interval and no observations in 70 to <80, making the 83-calorie beer visually isolated.

Step 7 - Compare beer types

Recommended beginner-friendly Excel approach

Create a small type list containing Regular, Light and Non-alcoholic. Use COUNTIF for counts and AVERAGEIF for means.

If you have Microsoft 365, use MEDIAN(FILTER(...)) and STDEV.S(FILTER(...)) for the median and standard deviation within each type. MINIFS and MAXIFS give the minimum and maximum. Then calculate CV as SD/Mean × 100.

Example for a type name in E2: =COUNTIF($B$2:$B$96,E2); =AVERAGEIF($B$2:$B$96,E2,$A$2:$A$96); =MEDIAN(FILTER($A$2:$A$96,$B$2:$B$96=E2)); =STDEV.S(FILTER($A$2:$A$96,$B$2:$B$96=E2)); =MINIFS($A$2:$A$96,$B$2:$B$96,E2); =MAXIFS($A$2:$A$96,$B$2:$B$96,E2).

If FILTER is unavailable, apply Data -> Filter to the Type column, copy each type's calorie values into a separate column, and apply MEDIAN and STDEV.S to each copied column.

Beer type	Count	Mean	Median	Std. dev.	CV	Min	Max
Regular	68	44.0147	43.0	7.2760	16.53%	26	83
Light	18	30.9444	30.5	5.8659	18.96%	19	43
Non-alcoholic	9	18.4444	16.0	5.7252	31.04%	14	30

Interpretation by type

Regular beers have the highest average calorie content at about 44.01 calories per 100 ml, followed by Light beers at about 30.94 and Non-alcoholic beers at about 18.44.

Regular beers have the largest absolute standard deviation (7.28), but Non-alcoholic beers have the largest coefficient of variation (31.04%). Therefore, Non-alcoholic beers are the most variable relative to their much lower mean.

The 83-calorie observation is a high extreme observation within the Regular category and is responsible for much of the strong positive skew and high kurtosis of Regular beers.

Step 8 - Create a box-and-whisker plot by type

28. Arrange the calorie values into three adjacent columns headed Non-alcoholic, Light and Regular. In Microsoft 365 you can use FILTER to spill each group into a column, for example =FILTER($A$2:$A$96,$B$2:$B$96="Regular").
29. Select the three columns including the headings.
30. Choose Insert -> Insert Statistic Chart -> Box & Whisker.
31. Give the chart the title Calories per 100 ml by Beer Type and label the vertical axis Calories per 100 ml.
32. Use the plot to compare the medians, IQRs, overall spread and unusual observations.

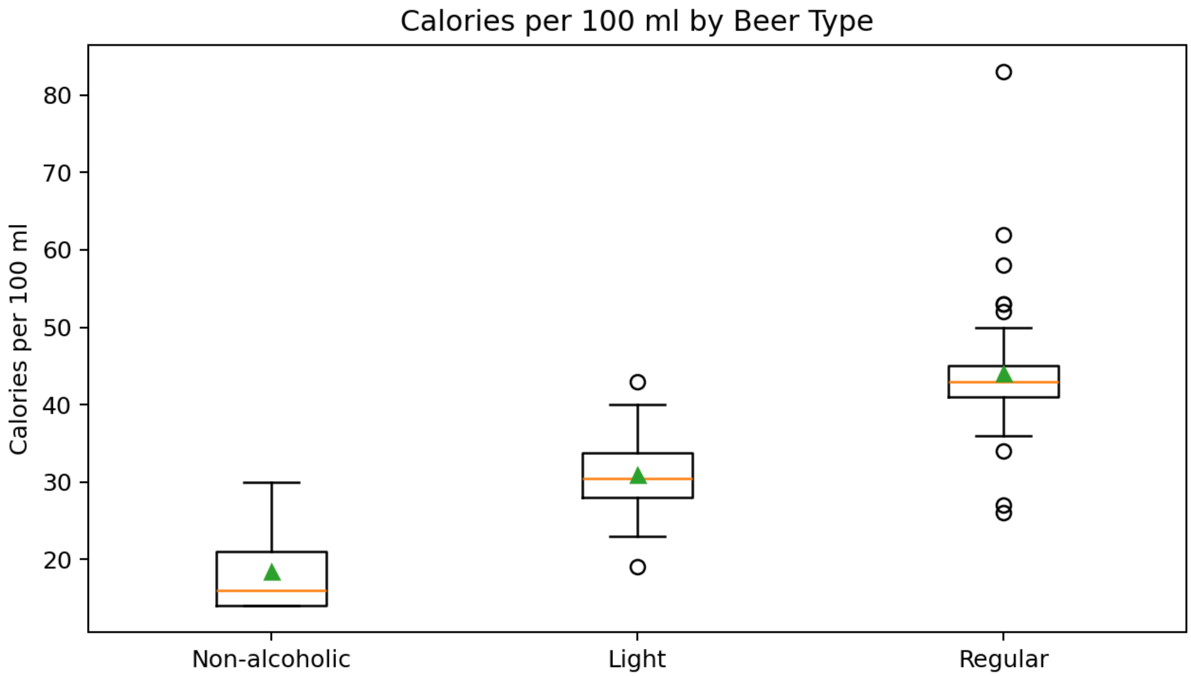

Figure 4: Box-and-whisker comparison of calorie content by beer type

Boxplot interpretation

The medians increase clearly from Non-alcoholic (16) to Light (30.5) to Regular (43). Regular beers are centred at substantially higher calorie levels.

The Regular category contains the conspicuous 83-calorie high observation. Because boxplot outlier rules depend on the quartile calculation convention, the exact number of plotted mild outliers can vary slightly across Excel settings; the substantive conclusion about the unusually high 83-calorie beer is unchanged.

Step 9 - Final written interpretation

Model answer

Across all 95 beers, the average calorie content is approximately 39.12 calories per 100 ml, with a median of 42 and a mode of 43. The sample standard deviation is 10.85 calories and the coefficient of variation is 27.75%. More than half of the sample (54.74%) lies in the 40 to <50 calorie interval.

The combined sample has near-zero skewness, but this should not be interpreted to mean that every beer type is symmetric. The mixture of three distinct product types masks important subgroup differences. Regular beers have the highest mean calorie content (44.01), Light beers are lower (30.94), and Non-alcoholic beers are much lower (18.44).

Relative variability is greatest for Non-alcoholic beers because its coefficient of variation is about 31.04%, compared with 18.96% for Light and 16.53% for Regular. The 83-calorie Regular beer has a Z-score of about 4.04 and is an extreme high outlier under the Lecture 3 Z-score rule.

A good business conclusion therefore uses both overall and type-specific summaries: the overall mean alone hides the large differences among product categories.

Checking checklist

- 95 calorie observations and no missing values confirmed.
- Complete ToolPak descriptive-statistics table produced.
- Q1, Q3, IQR and CV calculated with formulas.
- Z-score column completed and extreme outlier(s) identified.
- Frequency table totals 95 and percentages total 100%.
- Histogram has touching bars, title and axis labels.
- Type comparison table is complete and correctly formatted.
- Box-and-whisker plot compares all three beer types.
- Written interpretation discusses centre, spread, shape, outliers and differences by type.

PART 3 - Pearson MyLab

1. Please ensure that you have finished the MyLab Quiz 1 which is due yesterday!

Note: If you are having issues obtaining access to MyLab, re-read the instructions available in the Welcome topic. If you are still unsuccessful, email the teaching team explaining the issue. Homework and Quiz completion are important because the final e-test uses MyLab.