**Homework week 3**

**Question 1**

Below you find the cost (in cents) per 1-ounce serving for a sample of 13 chocolate chip cookies. The data are as follows:

54 22 25 23 36 43 7 43 25 47 24 45 44

1. **Determine the range**

54-7 = **47**

1. **Compute the first quartile (Q1), the third quartile (Q3) and the IQR**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Value | 7 | 22 | 23 | 24 | 25 | 25 | 36 | 43 | 43 | 44 | 45 | 47 | 54 |
| Position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |

n = 13

so the 7th value is the median (Q2) = 36

Q1 is between 3rd and 4th value = 23.5

Q3 is between 10th and 11th value = 44.5

IQR = 44.5 – 23.5 = **21**

1. **List the 5-number summary (minimum=L, Q1, median, Q3 and maximum=H)**

Minimum: 7

Q1: 23.5

Median (Q2): 36

Q3: 44.5

Maximum: 54

1. **Construct a boxplot and describe whether it is**

**skewed or symmetric**

The box consists of two equal parts which does not indicate skewness

The whiskers do not have the same length: this means skewness. Because the lower whisker is much longer the upper whisker, the distribution is skewed to the left.

1. **What can you conclude about the cost (in cents) per 1-ounce serving?**

The costs lie between 7 cents and 54 cents.

The middle 50% of the values lie between 23.5 cents and 44.5 cents

There is one chocolate chip cookie with very low cost per ounce compared to the rest; only 7 cents.

**Question 2**

Below you see the number of packages delivered by DHL for the first 8 days in February and the first 8 days in July.

**February**: 345, 213, 630, 530, 412, 190, 385, 410

**July**: 420, 109, 264, 283, 362, 526, 632, 610

1. Mention two disadvantages of using the range

Sensitive to extreme outliers and ignores the way the data are distributed

1. Calculate the range for February and July

February: 630-190 = 440

July: 632 - 109 = 532

1. Calculate the IQR for February and July

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Feb | 190 | 213 | 345 | 385 | 410 | 412 | 530 | 530 |
| Jul | 109 | 264 | 283 | 362 | 420 | 526 | 610 | 632 |
| position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

Split the data in half: 2 groups of 4

Q1 = the average of value 2 and 3, Q3 is the average of value 6 and 7.

February:

* Q1 = 279 and Q3 = 471
* IQR = 471-279 = 192

July:

* Q1 = 273.5 and Q3 = 568
* IQR = 568-273.5 = 294.5

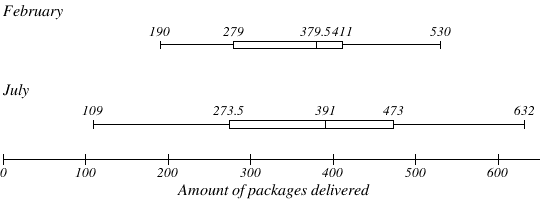
1. What can you conclude regarding the calculation at question c?

There is more variation in the amount of packages delivered in July compared to February

1. Draw the boxplot for February and July and use 1 x-axis

February: min = 190, Q1 = 279, median = 379.5, Q3 = 411, max = 530

July: min = 109, Q1 = 273.5, median = 391, Q3 = 473, max = 632



**Boxplots contain errors due to previous miscalculated Q3 values. Boxplots therefore only show an impression.**

1. Draw a conclusion based on the boxplot you have created.

For example:

There are no extreme outliers

There is more variation (spread) in the packages delivered in July (340) compared to February (523)

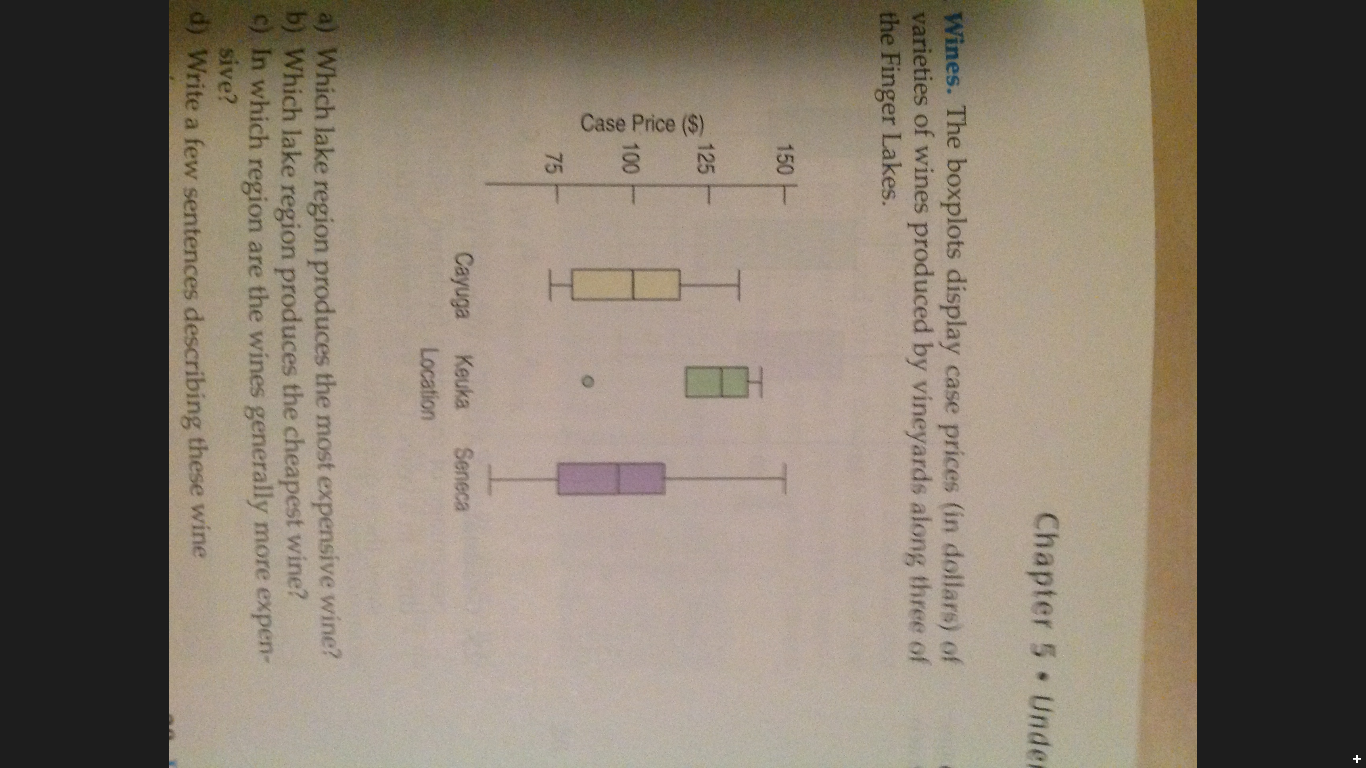
In July, the middle 50%of the values comprise a deliverance between 273.5 and 473

Please note that when drawing a conclusion, use exact numbers from the boxplots

The data of February is slightly skewed to the right, the data for July looks almost symmetric

**Question 3**

The boxplots below display case prices (in dollars) of varieties of wines produced by vineyards along three of the Finger Lakes.



1. **Which lake region produces the most expensive wine?**

Seneca

1. **Which lake region produces the cheapest wine?**

Seneca

1. **In which wine region are the wines generally more expensive?**

Keuka Location

1. **Which lake region has the largest range?**

Seneca

1. **Which lake region has the smallest IQR?**

Keuka Location

1. **Write a few sentences describing these wine prices.**

The wine prices in the 3 regions range from approximately 50 dollars to 150 dollars.

In Cayuga 50% of the wine prices are below 100 dollar. There are no extreme expensive wines.

In Keuka location the wines are in general most expensive. The prices are closer together compared to the other two regions. There is one relatively cheap wine of around 85 dollars, the rest of the wine prices are between 120 and 140 dollars.

The prices of Seneca vary the most. The cheapest wine is 50 dollars and the most expensive wine is 150 dollars. Seneca has the cheapest and the most expensive wine.

**Question 4**

Two rangers in South-Africa (Elvis and Abedi) count the number of lions they have spotted per day. Elvis counts the lions in the Kruger Park and Abedi counts the lions in Umfolosi National Park.

Elvis: 12 14 12 12 14 17 18 9 10 11

Abedi: 25 4 12 13 29 11 15 20 0 11

1. **Compute the mean, median and mode**

**Mean:**

Elvis: 129/10 = **12.9**

Abedi:140/10 = **14**

**Median Elvis:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Value | 9 | 10 | 11 | 12 | 12 | 12 | 14 | 14 | 17 | 18 |
| Position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

n = 10

The median is the average of the 5th and 6th value. The median is **12**.

**Median Abedi:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Value | 0 | 4 | 11 | 11 | 12 | 13 | 15 | 20 | 25 | 29 |
| Position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

The median is the average of the 5th and 6th value, which is **12.5**

**Mode:**

The mode for Elvis 12. The value occurs 3 times, which makes the distribution unimodal

The mode for Abedi is 11. The value occurs 2 times, which makes the distribution unimodal