Lab assignment 3
Statistics for EMCL students
solutions

Important note: Please observe that I always give the exact answer to the question, and nothing more, except of some explanatory notes. If you give me too much information, I don't know if you know which information is *the* answer to the question.

B.

*** 1. Copy the table in your report.**

MLU

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	2	10,0	10,0	10,0
	4	8	40,0	40,0	50,0
	5	2	10,0	10,0	60,0
	6	1	5,0	5,0	65,0
	8	4	20,0	20,0	85,0
	9	1	5,0	5,0	90,0
	10	1	5,0	5,0	95,0
	11	1	5,0	5,0	100,0
	Total	20	100,0	100,0	

*** 2. How many measurements (data) do you have?**

n=20

*** 3. Which MLU is the second most frequent?**

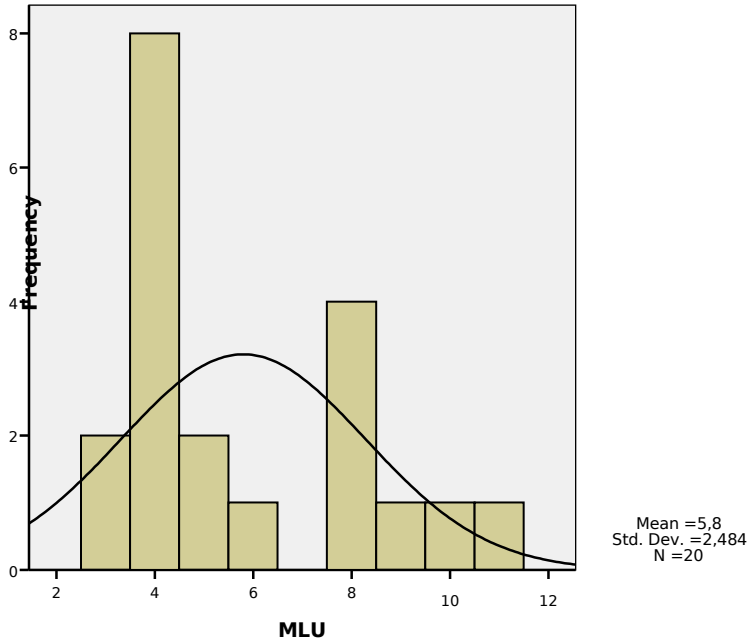
Second most frequent value is 8 (namely, its frequency is 4).

*** 4. How often does the highest value of MLU occur?**

Its frequency is 1 (because the highest MLU value is 11).

C.

*** 5. Copy this second graph to your report.**



*** 6. What does the vertical axis display: numbers or percentages?**

Vertical axis: numbers (of occurrences).

*** 7. What is the highest value and what is the lowest value of the variable?**

Highest value: 11, lowest value: 3.

*** 8. How many peaks are there?**

There are 2 peaks.

*** 9. There is a gap is the graph. At what value can this gap be found? What does this observation mean? Would you expect to find this gap if you had many more data?**

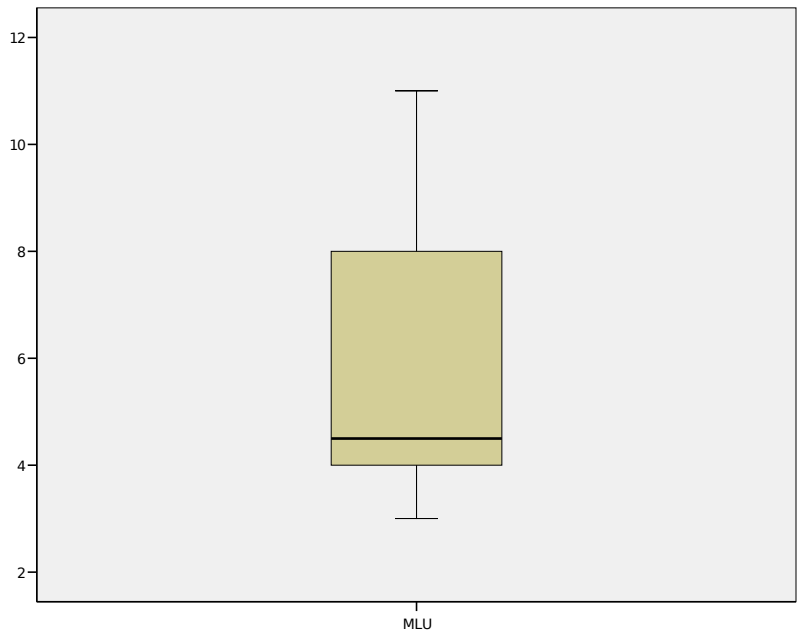
Gap is at value 7. It means no observation has value MLU = 7.
If much more data, then probably there would be no such gap, because most probably this is only a **random gap**. Namely, there is no reason to have a bimodal distribution with two peaks, and have a **systematic gap** between them. Value 6 has frequency 1, value 8 has a the highest frequency (4), so it would be surprising to have the mode of the distribution just after a gap. But you never know, unless you have a theory of the phenomenon at hand that you are very confident about...

*** 10. Is this distribution approximately Normal?**

No, it is quite far away from a Normal distribution: the best Normal curve that can be fitted to the data is still very different from the observable distribution. Obviously, as you have relatively few observations, so it may still be the case that the data are distributed in a Normal way but you are unlucky. So far, the data do not support an argument for a Normal distribution.

D.

*** 11. Copy this boxplot to your report.**



*** 12. Which is the lowest and highest value according to the boxplot?**

Lowest: 3, highest: 11 (the two ends of the whiskers).

*** 13. Which is approximately the median according to the boxplot?**

Median: 4.5 (the middle line within the box).

*** 14. How many percentages of the data are outside of the box?**

2 data below, 7 data above, that is 9 data of the 20.

Explanation: In general, at most 50 % of data outside of the box: at most 25% below (lower than Q1) and at most 25% above (higher than Q3). If none of the observations (data points) lie exactly on the first and third quartiles, then you have exactly 50% of your observations outside of the box.

*** 15. Which data are outside of the ?whiskers? of the boxplot?**

0%.

Explanation: there are no outliers according to the $1.5 \times IQR$ rule in this data set. But try out what happens if you added a 21st piece of data whose value is 25. Created again the boxplot, and you will see that this outlier will be located outside the whisker. For more explanation, see Chapter 1 of M&M on "modified boxplots". In fact, we now learn that what calls SPSS a boxplot is what M&M calls a "modified boxplot".

See also: http://www.mathwords.com/b/boxplot.htm.

E.

*** 16. Copy this table to your report.**

Statistics

MLU

N	Valid	20
	Missing	0
Mean		5,80
Median		4,50
Mode		4

*** 17. Suppose you make an error during data entry: you type 80 instead of 8. Which of these values will change, and which will not? (Why? How does M&M call this feature of a statistical measure?)**

Mode, median does not change because they are *resistant measures* (M&M, p. 32): as long as the most frequent value remains the same, and as long as a data point does not move from the upper half of the observations to the lower half (or vice-versa), the remain the same. Mean changes radically, because it is dependent on the exact value of each data point.

*** 18. The median of MLU is lower than its mean. This is because the histogram is skewed to the ? (left or right?), and it has a longer tail to the ? (left or right?).**

Skewed to the *right*, longer tail to the *right*.

E.

*** 19. Report the SD, the range and the IQR.**

SD = 2.484 range = 8 IQR = Q3 – Q1 = 4

Explanation:

range = max – min = 8

Q1 = 25^{th} percentile = 4

Q3 = 75^{th} percentile = 8

*** 20. If the range is seen as the width of the histogram, then how many SD is the width of this histogram?**

Range / SD = approx. 3.22 (using the Calculator of Windows).