Quizzes : Module 2 Part 1

Statistics and Visualization

In the LAMS Sequence, you have learned the theory behind this module. It is also expected that you have attempted the quizzes embedded within the LAMS Sequence, and have used the “unlimited attempts” opportunity to score 100%. Here are the quiz questions, consolidated with their answers and corresponding feedback. This is for your after-LAMS revision.

Question 1

Suppose the Male vs Female ratio in a specific class of NTU is 70% vs 30%, whereas you know that the Male vs Female ratio in Singapore is 960 against 1000. Do you think that the selection process for admission was biased towards male applicants?

Answer Choice	Verdict	Explanation
I can't be sure, unless I get the data for the applications, and know the gender ratio within applicants.	Correct	This is more accurate. Once you know the applicants' profile and ratio, you can judge the selection process better.
Possibly, as gender ratio in the specific class is unexpectedly different from that in the country.	Wrong	Think again. Was the proportion of applicants to the specific program exactly the same as the ratio in Singapore? If not, then you can't conclude this.
No. As the gender ratio in the specific class is same as the gender ratio in that specific field of study across Singapore.	Wrong	Yes, you are thinking in the right track. But is the field of study everything? Think again. Was the proportion of applicants to the specific program exactly the same as the ratio in the field of study? If not, then you can't conclude this.
No. As the gender ratio in the specific class is same as the gender ratio in this field across all JCs and Polys.	Wrong	Yes, you are thinking in the right track. But is the JC or Poly background everything? Think again. Was the proportion of applicants to the specific program exactly the same as the ratio in the JCs or Polys? If not, then you can't conclude this.

Reference Module 2 Topic 1 : Uni-Variate Statistics No specific slide. The overall concept matters.

Question 2

Suppose the average household income of Singapore is SGD 12,000 per month, while the average household income of USA is SGD 15,000 per month. What can you conclude?

Answer Choice	Verdict	Explanation
Can't compare the household income distributions of the two countries with just this information.	Correct	True. You do not have enough data yet. You will need Median, Standard Deviation and Number of Households to know more.
The top 50% households in USA earn more per month than the top 50% households in Singapore.	Wrong	Do you know the Median in either case? If not, then you DO NOT know about top 50% of either country. Think again.
Household income inequality in USA is significantly more than the income inequality in Singapore.	Wrong	Do you know the Standard Deviation in either case? If not, then you DO NOT know about income inequality or spread of either country. Think again.
Total household income across USA is higher than the total household income across Singapore.	Wrong	Do you know the total number of households in either case? If not, then you DO NOT know about total household income of either country. Think again.

Reference Module 2 Topic 1 : Uni-Variate Statistics Slide 6

Question 3

Suppose you have a dataset X, with n numeric values in total. Which of the following values will be the largest?

Answer Choice	Verdict	Explanation
Impossible to tell. Depends on the actual data.	Correct	True. You can't tell without looking at the data. In fact, you can come up with example datasets where each of the other items will be the greatest.
Mean	Wrong	Think again. Can you say for sure, without looking at the actual data? Can't mean be 0 with a high standard deviation?
Mean Absolute Deviation	Wrong	Think again. Can you say for sure, without looking at the actual data? What if all the values are the same?
Standard Deviation	Wrong	Think again. Can you say for sure, without looking at the actual data? What if all the values are the same?

Reference Module 2 Topic 1 : Uni-Variate Statistics Slide 6 and Slide 7

Question 4

Suppose the average household income of Singapore is SGD 12,000 per month, and the standard deviation of the household income is SGD 3,000 p/m. What can you conclude?

Answer Choice	Verdict	Explanation
None of the other statements are valid under all cases, as it depends on various other factors.	Correct	True. You DO NOT know enough about the distribution yet. Only Mean and Standard Deviation is not enough, unless it is truly Normal Distribution. And you can't assume that.
More than 50% of the households in Singapore earn between SGD 9,000 to SGD 15,000 per month.	Wrong	You are assuming a Normal Distribution, implicitly. Can you assume that? Think again.
Maximum household income inequality in Singapore is no more than SGD 6,000 per month.	Wrong	Not true. You just know that SGD 6,000 per month is twice the standard deviation. You don't know the minimum or maximum.
Minimum and maximum possible household income in Singapore is SGD 3,000 and SGD 21,000.	Wrong	Not true. You can't say this without knowing the actual minimum and maximum household income.

Reference Module 2 Topic 1 : Uni-Variate Statistics Slide 6 and Slide 7

Question 5

Suppose that the mean score of your class is 75, the median score of your class is 77, the standard deviation of the scores is 5, and your own score is 81. What can you conclude?

Answer Choice	Verdict	Explanation
You definitely have a score better than 50% (or more) students in your class.	Correct	True, as you are above the Median for sure.
Definitely more than 50% students in the class scored above-average marks.	Correct	True, as the Median is higher than the Mean.
You definitely scored better than 84% (or more) students in your class.	Wrong	Not sure, unless you know that it is a Normal Distribution (you can't assume that).
Your score is above-average, but may be in the lower 50% students in the class.	Wrong	Can't be, as you scored above the Median.

Reference Module 2 Topic 1 : Uni-Variate Statistics Slide 6 and Slide 7 and Slide 8

Question 6

Suppose the median household income in Singapore is SGD 9,000, and the quartiles are Q1 = SGD 4,500 and Q3 = SGD 10,500. What can you infer from this data?

Answer Choice	Verdict	Explanation
70% or more of the households in Singapore earn below SGD 10,500 per month.	Correct	True, as 75% of the data must lie below the third quartile.
25% or more households in Singapore earn more than SGD 10,000 per month.	Correct	True, as 25% of the data must lie above the third quartile.
70% or more of the households in Singapore earn between SGD 4,500 to SGD 10,500 per month.	Wrong	Not true, as only 50% of the data lies within the first and third quartiles.
None of the households in Singapore earn more than SGD 1 Million per month.	Wrong	You don't know that. There may always be outliers, and quartiles do not tell you anything about them.

Reference Module 2 Topic 1 : Uni-Variate Statistics Slide 8 and Slide 9

Question 7

Suppose the median household income in Singapore is SGD 9,000, and the quartiles are Q1 = SGD 4,500 and Q3 = SGD 10,500. What can you infer about the outliers?

Answer Choice	Verdict	Explanation
Household income above SGD 20,000 may be considered as outliers (abnormally high) in this data.	Correct	That's quite far indeed. It's higher than (Q3 + 1.5 * IQR), where Inter-Quartile Range IQR = Q3 – Q1. That may be considered outlier as per standard norms.
Household income above SGD 10,500 may be considered as outliers (abnormally high) in this data.	Wrong	Nope. There are of course 25% of houses above third quartile (Q3), and they are not all outliers.
Household income less than SGD 1,000 may be considered as outliers (abnormally low) in this data.	Wrong	Not quite. It's still within the Q1 to (Q1 – 1.5 * IQR) interval on the lower side, and can’t be termed outliers.
We can't say that any of the other answers are true unless we know the average household income.	Wrong	Some justification about outliers can be drawn from the quartile gaps. Think again.

Reference Module 2 Topic 2 : Uni-Variate Visualization Slide 5

Question 8

Suppose you want to know where you stand in terms of annual income in Singapore. Which one of the following statements would you support in this context? Select all that you think are right.

Answer Choice	Verdict	Explanation
Knowing the average income does not help at all, as the average may be affected by really high/low outliers.	Correct	Somewhat true. Outliers affect the average or the mean quite heavily, and hence it is not a robust indicator. However, can't say it "does not help at all".
It's good to know the median income in Singapore to judge which percentage of the demography I belong to.	Correct	True, as the median will definitely tell you if you are in the higher 50% or the lower 50% of the population. Nothing more though.
Histogram/KDE of income in Singapore is a much richer source of information compared to a quartile box plot.	Correct	No wonder -- the box plot only tells you which quartile you belong to. However, the histogram shows detailed distribution.
It's good to know the average income in Singapore to judge where I stand with respect to the population.	Wrong	The average or the mean is NOT a robust indicator -- it is highly influenced by high/low outliers.

Reference Module 2 Topic 2 : Uni-Variate Visualization No specific slide. The overall concept matters.