Extension Statistics Practice #2

You are a selector for the New Zealand Synchronised Swimming Federation.

Two teams have reached the qualifying mark, but only one of them may be selected to go to the Olympics.

You must select the best team.

The scores in competitions for the last year for the two teams are:

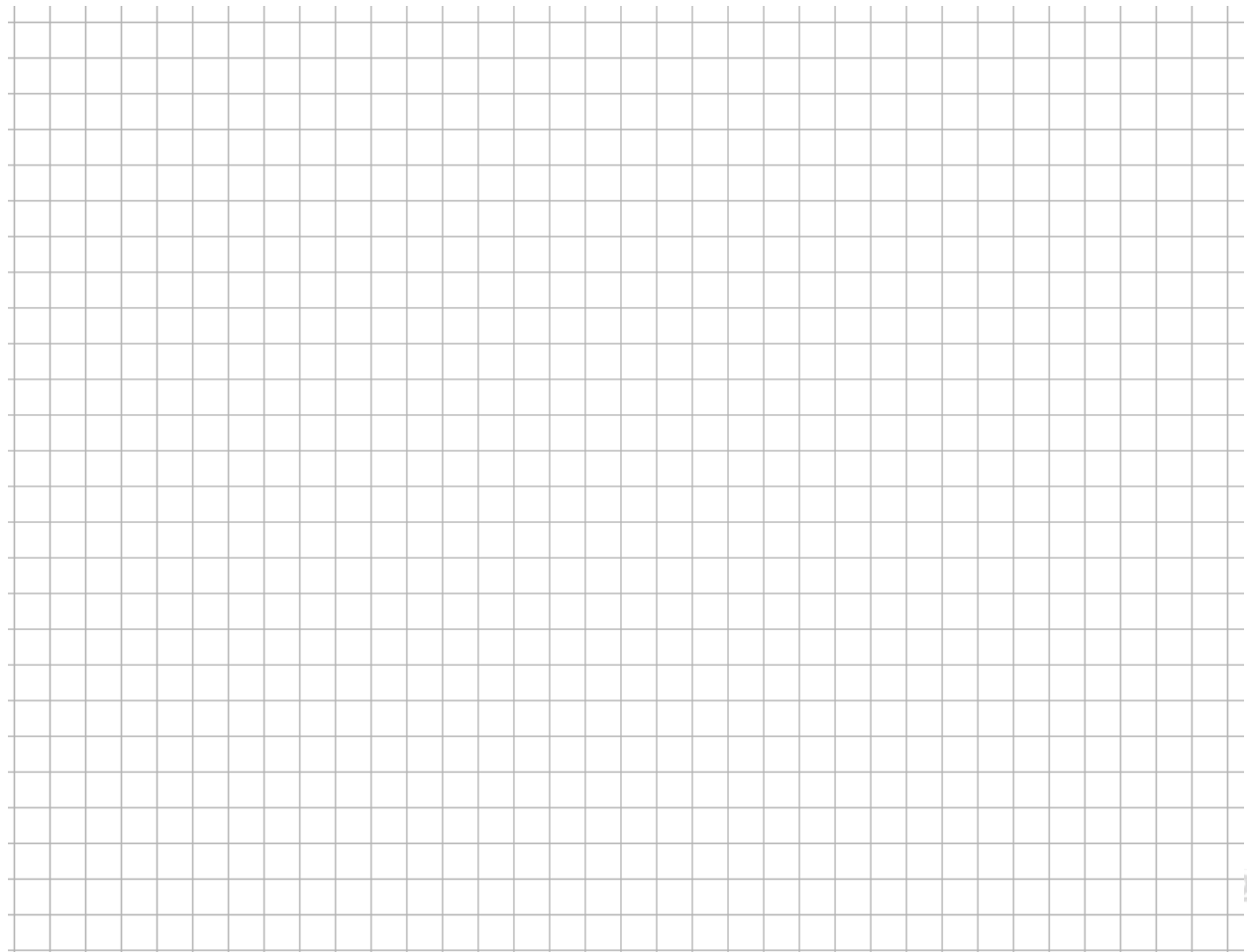| Event	Team 1	Team 2
1	35.4	27.8
2	32.1	30.2
3	34.5	30.1
4	29.7	dnc
5	33.6	33.3
6	38.1	32.8
7	31.9	dnc
8	33.5	34.2
9	35.3	33.9
10	33.8	35.2

dnc = did not compete

Scores are out of 40. Higher scores are better.

Task

You must prepare a report

- summarising the scores in some manner,

- analysing the scores, with at least one suitable graph to visually summarise the information,

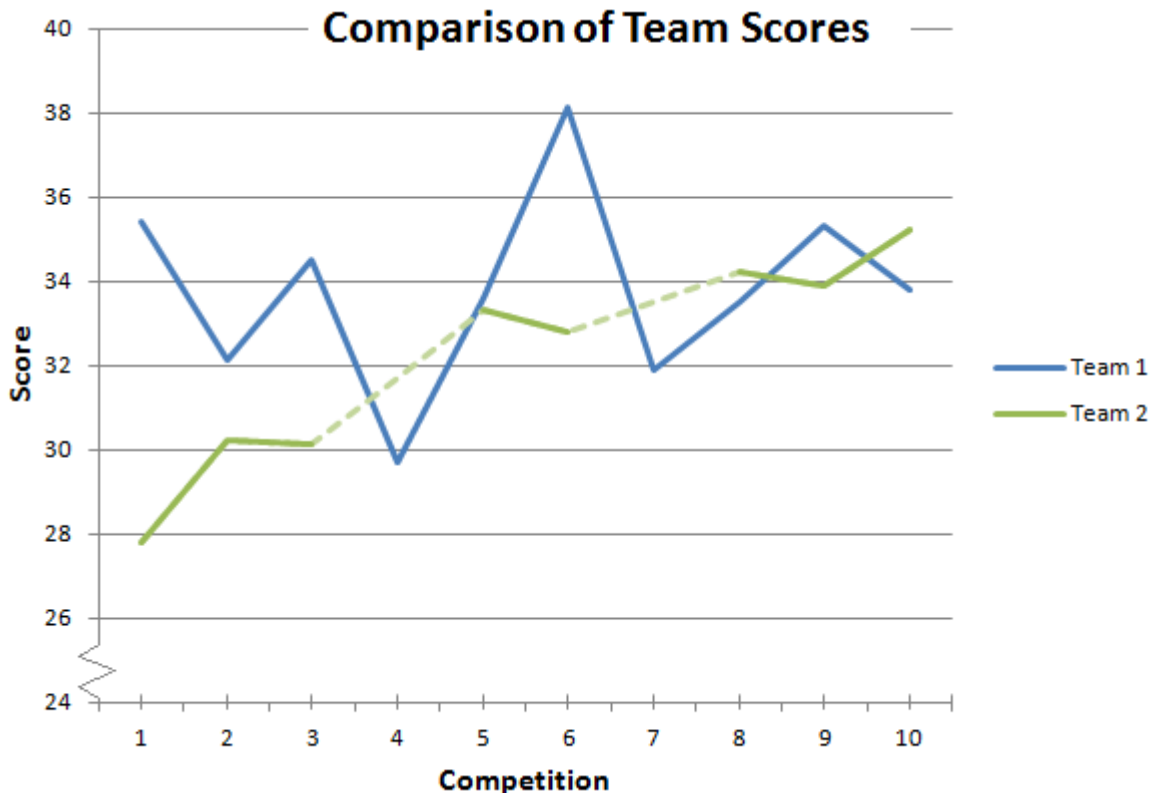- with a recommendation on who to send to the Olympics based on your analysis.
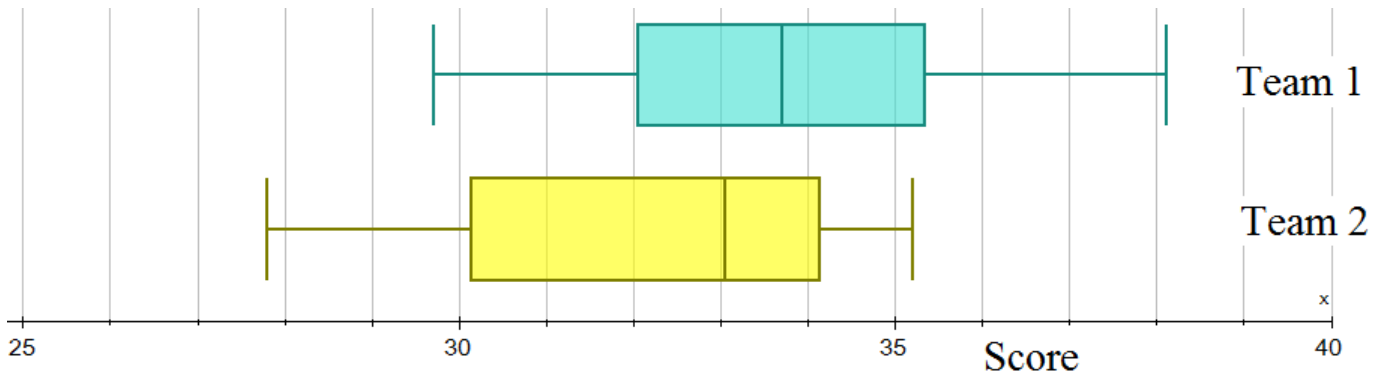
Model Answers: Extension Statistics Practice #2

Statistics calculated

Team 1	Mean = 33.79	Median = 33.7	Range = 8.4	IQR = 2.65
Team 2	Mean = 32.19	Median = 33.05	Range = 7.4	IQR = 3.8

Scores over the Season

Comparison of Team Scores

Note: you cannot put a value of 0 in for the "dnc" results. Statistics must be calculated only on the basis of the actual data (so the mean for Team 2 is divided by 8, as there are only 8 valid scores).

For a time series graph you can link across missing points with dotted lines, or leave a gap, but you must not go to zero with data that is missing. That gives entirely the wrong shape to the data.

Example Analysis #1

I am going to base my decision on which team is likely to get the better score in the next competitions.

The data shows that while it seems that Team 1 was much more erratic than Team 2 (their range is larger at 8.4 compared to 7.4) the time series graph shows that Team 2 is actually very consistently getting better (positive trend), whereas Team 1 has no pattern over time. The large range and inter-quartile range are meaningless in such a situation. The medians for both are within the IQR for the other, so we cannot be confident that they are statistically significant.

If the current trend for Team 2 to keep improving their scores keeps going then they will start to outscore Team 1 fairly consistently. That means they are most likely to get the highest score in the next competition.

I would recommend Team 2, based on their pattern of improvement.

Example Analysis #2

I am going to base my decision on which team is likely to get a very high score, since to win overseas a very high score will be required.

Team 1 is more erratic than Team 2 (their range is larger at 8.4 compared to 7.4) but their top quarter of values are higher than the highest value by Team 2 as can be seen on the box-and-whisker graph.

Team 1 has a better chance to get a really good score and win although they also have a much better chance of doing badly as well.

The difference in the medians is not enough to be statistically significant as both lie well within the IQRs for the other team. Therefore I do not feel that I can base my decision on the slightly higher median for team 1, because it might just be the result of random sample variation.

I would recommend Team 1, as they are most likely to score a very high score.

_____

Notes for Students

It is not unusual for a problem to have more than one "correct" answer. However it is important that your answer:
- Is made on whatever basis that you decide (or are given) at the start of your analysis.
- Is defended using statistics (always mention the numbers).
- Does not ignore entirely the advantages of the other answer.

In a situation like this, for Merit and Excellence you could suggest improvements such as:
- You need more information about what a good or winning score is.
- You need information about why Team 1 is so erratic.
- Some indication if Team 2's trend of improving is likely to continue in the future.
- How soon is the overseas competition? Do Team 2 have time to improve more?

While this data can be analysed without using a time series line graph, to not do so would be to lose all chance of seeing the trend of Team 2. That would make it hard to get a good mark.

If the data given is over a period of time, then usually a line graph with an x-axis of time is the best starting point. Even not seeing a trend over time is worth commenting on. Only after that do you fall back onto the normal statistics (medians, IQRs, ranges, means).