Name

A teacher tries to assess her performance in class by giving the students a simple test.  
The students seemed to score well on a test out of 100.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 90 | 86 | 96 | 89 | 94 | 93 | 78 | 94 |
| 85 | 94 | 83 | 90 | 87 | 94 | 87 | 95 |
| 83 | 89 | 82 | 72 | 92 | 93 | 67 |  |
| 72 | 87 | 77 | 93 | 99 | 89 | 83 |  |
| 95 | 95 | 97 | 88 | 79 | 84 | 93 |  |

She reported   
My students scored well with a mean 87.676 with  = 7.494, so my lesson was successful.

Comment: Did she make a good report? Is there anything wrong with it?

*(You do NOT need to check the histogram – except for your own practice!)*

1. The distribution of scores is skewed severely! This should have been noted!
2. Because of the skew, the standard deviation is NOT valid.
   1. Result should be described in another way, i.e. using a skewness metric
3. Too many significant digits for sample of 37. Report should be mean = 88 ± ??  
   ?? = range or skewness metric, eg kurtosis

**Reproducibility** is an important scientific principle. Suggest two **hypotheses** on which you could base further tests to verify the teacher’s claim that her methods were effective. Descibe a **simple** test to verify these hypotheses (Short: one or two sentences each!).

Background: Usually we would expect a normal distribution of general abilities in a class. Therefore a normal distribution of scores.

H1: If the test was made more difficult, then we would expect a normal distribution of results.

Method: Add some more difficult questions to the test.

H2: If another teacher set the test, then it would not accidentally reflect the teacher’s bias.

Method: Ask another teacher to set some (or all) of the questions.

H3: If the test was longer and more comprehensive, then it might reflect overall retention of the lesson.

H4: ***Your ideas?? Be imaginative ☺***