Top Ten List of Things to Know to Pass Introductory Statistics

10. *Good results cannot be coaxed from bad data by doing more calculations.*

9. *A sample selected so that each unit in the population has an equal chance of being in the sample is not necessarily a simple random sample.*

8. *A hypothesis test does not <u>prove</u> anything.*

7. *In many situations, variances add. Standard deviations do not add.*

6. *Correlation does not imply cause and effect.*

5. *When summarizing quantitative data, measures of spread are at least as important as measures of center.*[*]

4. *Know (and check!) the conditions necessary to ensure the validity of the statistical procedure you want to use. Know what can go wrong.*

3. *Beware of outliers*.

2. *Data are numbers with a context.*

1. ***Starting point for all data analysis: make a picture!***

[*] Students tend to focus excessively on the mean or median and ignore variation. You do so at your own peril. Measures of center are abstractions that may not apply to any data point or individual. Variation is the reality; focus more attention on it. As an example, consider the following:

EXAMPLE:
You take 2 exams in a statistics course. Your grades are:
exam 1, 92; exam 2, 90.

The class mean and standard deviation for the exams are as follows:
exam 1: mean = 80, standard deviation = 10
exam 2: mean = 80, standard deviation = 8.

Most people would say that the 92 on exam 1 is a better grade than the 90 on exam 2. After all, the means for both exams are the same (80) so 92 must be better than 90. IT ISN'T!! The 90 on exam 2 is the better grade. Let's see why.

The score of 92 on exam 1 is 1.20 standard deviations above the exam 1 mean of 80: the z-score for the exam 1 grade of 92 is (92-80)/10 = 12/10 = 1.20.

The score of 90 on exam 2 is 1.25 standard deviations above the exam 2 mean of 80: the z-score for the exam 2 grade of 90 is (90-80)/8 = 10/8 = 1.25

If each exam was graded on the curve, the grade of 90 on exam 2 would be given a higher grade than the 92 on exam 1 since 90 is more standard deviations above its respective mean. Note that 90 is better than 92 even though exams 1 and 2 have the same mean.

In the example above the mean was 80 but there may not have been any exam with a score of 80. The mean of a set of numbers does not have to be one of the numbers.

What would you guess is the mean number of arms on a student at NC State? If you guessed 2, you're wrong. All it takes is 1 student who is missing an arm and the mean number of arms on the 28,000 students is less than 2 (I saw a student the other day who was missing an arm). So the mean number of arms on students at NCSU is 1.99999, say. So what? Would anyone make a sweater with 1.99999 arms? Of course not. If 5 students are missing one or both arms, then 99.98% of the students have more than the mean number of arms; so in this case the mean is not a very good measure of center.

The variation of the number of arms on students (and people in general) is small so sweater makers make sweaters with 2 arms and that's it. Not so with jeans. If Levi Strauss made jeans with only 1 waist size (say the mean waist size) how long do you think they would be in business? There may not even be a person alive whose waist is the mean waist size. Levis Strauss takes variation into account and makes jeans with many different waist sizes.