

MATH 156.6 — TEST I*PLEASE READ THESE INSTRUCTIONS:*

Problems have varying point values, which are specified next to each problem number. Move quickly through the test, but try to be as accurate as you can. Do as many problems as you like; points over 100 will count as extra credit.

m If you want credit for an answer, *you must explain what you are doing* – this means defining your terms and explaining **each step** in your work.

You may use a graphing/statistical calculator. You may not use your book or notes or a formula card.

PLEASE TURN OFF YOUR CELL PHONE.

[Please do not write in the boxes below:]

1a :	/8
------	----

1b :	/8
------	----

1c :	/5
------	----

1d :	/8
------	----

1e :	/8
------	----

1f :	/8
------	----

1g :	/4
------	----

2a :	/8
------	----

2b :	/8
------	----

2c :	/6
------	----

2d :	/4
------	----

3a :	/4
------	----

3b :	/4
------	----

3c :	/4
------	----

3d :	/4
------	----

3e :	/4
------	----

4a :	/6
------	----

4b :	/8
------	----

5a :	/6
------	----

5b :	/6
------	----

5c :	/6
------	----

6a :	/6
------	----

6b :	/8
------	----

Σ :

- [49] **1** Here are the numbers of home runs that Babe Ruth hit in his 15 years with the New York Yankees, 1920 to 1934:

54 59 35 41 46 25 47 60 54 46 49 46 41 34 22

- [8] **(a)** Make a stemplot for these data.
- [8] **(b)** Make a histogram these data.
- [5] **(c)** Is the distribution symmetric, clearly skewed (in which direction, then?) or neither? Why?
- [8] **(d)** About how many home runs did Ruth hit in a typical year – so, where is the **center** of the distribution of these home run data? Answer this question with at least two numbers and explain the differences between them.

[8] (e) Was Ruth a “consistent” home run hitter, or was there a lot of variation in how many he hit? Give at least two numerical measures of his variability as a home run hitter, and explain the difference in these measures.

[8] (f) Make a box-and-whisker plot of these data. Explain how each part of the plot is determined and what it means.

[4] (g) Was the famous 1927 season with 60 runs an outlier? Why or why not?

[26] **2** A strange fellow in Australia collected some data over a period of days, by measuring the weight (in grams) of the bar of soap in his shower. Here is the data: [Note that he missed some days.]

Day:	1	2	5	6	7	8	9	10	12	13	16	18	19	20	21
Weight:	124	121	103	96	90	84	78	71	58	50	27	16	12	8	6

[8] (a) Make a (rough sketch of a) scatterplot of this data. (But label it very precisely!)

[8] **(b)** Compute the correlation of the two variables. What does its sign tell us? Its magnitude? (Phrase your answers to these last questions both in purely mathematical terms and also in practical terms about this man and his soap.)

[6] **(c)** Compute the coefficients of (and write down explicitly) the equation of the least squares regression line. Plot this regression line on top of your scatterplot.

[4] **(d)** What would you predict would have been the weight on day 17, one of the days he forgot to measure?

[20] **3** Give examples of small data sets (say 3 or 5 or 7 values) for which:

[4] **(a)** the mean and the standard deviation are both 0;

[4] **(b)** the mean is 0 and the standard deviation is greater than 10;

[4] (c) the mean is 0 and the median is 100;

[4] (d) the median is 0 and the mean is 100; and

[4] (e) the mean is larger than the third quartile.

[14] **4** Some work with Normal distributions:

[6] (a) The NCAA requires Division I athletes to have at least a combined SAT score of 820 (so, the cut-off is 820). In 2002, SAT scores were approximately $N(1020, 207)$. What percent of all students were not NCAA eligible for Division I sports?

[8] (b) Say the NCAA wanted to make a new, super academic Division which would only allow the top 2.5% of SAT scorers to be eligible. What should they have set the SAT cut-off to be in order to achieve this in 2002? [*Hint: use the “68-95-99.7 Rule”.*]

[18] **5** A medical journal describes the results from a certain study, beginning with the phrase “The Physicians’ Health Study is a randomized, double-blind, placebo-controlled trial designed to determine whether....” Explain to a doctor who knows no statistics

- **what** each of the following words means;
- **why** it is important; and
- **how** it could have been achieved:

[6] (a) “randomized”;

[6] **(b)** “double-blind”; and

[6] **(c)** “placebo-controlled”.

- [14] **6** Suppose a committee of math professors is interested in the average number of homework problems per chapter in a new statistics textbook, but they are too lazy to count them all. So perhaps they decide simply to take a Simple Random Sample of five chapters and to see the average for this sample.
- [6] **(a)** What is a “Simple Random Sample”, in general?

- [8] **(b)** The chapters are numbered 1 through 29. Explicitly find an SRS of size 5 from among all these chapters [so: tell me the five chapter numbers you will include in your sample, describing exactly how you chose those numbers]; use, if you like, the following line from a Table of Random Digits:

1 7 1 3 9 2 7 8 3 8 1 9 1 3 9 8 2 0 3 1 4 6 1 4 3