

MATH 156.3 — TEST II*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.

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 : | /11 |
| 1b : | /11 |
| 1c : | /8 |
| 1d : | /10 |
| 1e : | /11 |
| 1f : | /8 |
| 1g : | /8 |
| 1h : | /5 |
| 1i : | /12 |
| 2a : | /3 |
| 2b : | /6 |
| 2c : | /8 |
| 3a : | /15 |
| 3b : | /9 |
| 3c : | /6 |
| 3d : | /9 |
| 3e : | /6 |
| Σ : | |

1

Here is a stemplot of the travel times (in minutes) to work of 15 randomly selected workers in North Carolina in 2004:

| | |
|---|-------------|
| 0 | 5 |
| 1 | 0 0 0 0 2 5 |
| 2 | 0 0 5 |
| 3 | 0 0 |
| 4 | 0 0 |
| 5 | |
| 6 | 5 |

[2+3×3] (a) What is the **mean** of this data?

What is the definition of the mean in general?

What qualitative feature of the data does the mean measure?

What are its strengths or weaknesses when compared with other measures of this same qualitative feature?

[2+3×3] (b) What is the **median** of this data?

What is the definition of the median in general?

What qualitative feature of the data does the median measure?

What are its strengths or weaknesses when compared with other measures of this same qualitative feature?

[3+5] (c) What are the quartiles **Q1** and **Q3** for this data (also called **Ql** and **Qu**)? What are the definitions of these quartiles in general?

[5+5] (d) What is the **five-number summary** of this data? Make a graphical representation of this five-number summary. What is this kind of plot called?

[2+3×3] (e) What is the **IQR** of this data? What is the definition of the IQR in general? What qualitative feature of the data does the IQR measure? What are its strengths or weaknesses when compared with other measures of this same qualitative feature?

- [2+2×3] (f) What is the **standard deviation** of this data?
What qualitative feature of the data does the standard deviation measure?
What are its strengths or weaknesses when compared with other measures of this same qualitative feature?

- [2+3×2] (g) What is the **range** of this data?
What is the definition of the range in general?
What qualitative feature of the data does the range measure?
What are its strengths or weaknesses when compared with other measures of this same qualitative feature?

- [2+3] (h) What is the **mode** of this data?
What is the definition of the mode in general?

- [4+3+5] (i) Are there any outliers in this data? If so, what are they?
What is the general definition of outliers you are using?
Make a modified version of the plot you made above in (d) which shows clearly the presence or absence of any outliers you may have found.

- 2 Suppose that in a certain population of students, SAT scores are normally distributed (what our book calls “mound-shaped”; in the lay press it is often called a “bell curve”) with mean 1500 and standard deviation 240. Say a particular student Joe Garcia got a 1740 on the test.
- [3] (a) How many standard deviations above the mean did Joe score?

- [6] (b) What percentage of the population scored within the same distance from the mean as did Joe?

- [8] (c) In what range of scores did the middle 95% of the students' scores lie?

3

Here is some data from 2007 on various airlines, telling the percentage of their maintenance that they outsourced to other companies, and the percentage of the delays caused by their own operations (*i.e.*, not caused by weather, air traffic delays, or other factors beyond the airline's control):

| <u>Airline</u> | <u>Outsourced(%)</u> | <u>Delays(%)</u> |
|----------------|----------------------|------------------|
| Hawaiian | 93 | 60 |
| Alaska | 92 | 42 |
| USAir | 77 | 24 |
| AmWest | 76 | 39 |
| Continental | 69 | 20 |
| JetBlue | 68 | 18 |
| AirTran | 66 | 14 |
| United | 63 | 27 |
| American | 46 | 26 |
| ATA | 18 | 19 |

[15] (a) Sketch here a scatterplot of these data, using *percentage of delays* as the response variable.

[9] (b) Describe the association between these variables that you observe. [*Hint: comment on form, strength and direction.*]

[6] (c) Is it reasonable to conclude that there is a cause-and-effect relationship between these variables, based on these data? Explain fully.

[2+2+5] (d) What is the correlation coefficient between the given airline performance variables? Does its sign and magnitude agree with your discussion in (b)? What do the sign and magnitude of the correlation coefficient tell us in general about the association of two random variables?

[6] (e) What would you expect the correlation coefficient to become (generally – don't give a specific number) if there had been one more airline which outsourced about 30-40% of its maintenance and had delays of around 60-70% due to its own operations? Why?