

Problems to Review Descriptive Statistics

STAT 401F, Fall 2007

For practice only! Not to turn in! Solutions will be provided

1. Salaries of professional athletes receive a good deal of attention in the press. The 1990 salaries of the non-pitchers on the Chicago Cubs baseball team are listed below, units are in thousands of dollars (e.g., the value of 100 units below means 100,000 dollars).

100	100	111	114	165	210	225	225
230	575	1200	1900	2100	2100	2650	3300

- (a) Calculate the median and the mean for these data.
 - (b) Suppose the owner of the team says that a typical player earns close to a million dollars (1000 units in thousands of dollars). Which measure of center is the owner using? Would it be accurate to say that a typical player earns close to a million dollars? Explain.
 - (c) Calculate a five number summary for these data.
 - (d) If each player received an extra 200,000 dollars (or 200 units in thousands of dollars), how would the mean and median change? (**You don't have to change the data and recalculate the mean and median. Instead, think about what each measures and how that is affected by adding a constant.**)
 - (e) If each player received an extra 200,000 dollars (or 200 units in thousands of dollars), how would the standard deviation change? (**You don't have to change the data and calculate the standard deviation. Instead, think about what it measures (i.e., spread) and how that is affected by adding a constant.**)
2. The sample average age of 5 persons in a room is 30 years. A 36-year-old person walks into the room. What is now the average age of the persons in the room? Suppose the median age is 30 years and a 36-year-old person enters the room. Can you find the new median age from this information?
 3. This is a sample variance context. You must give a list of two numbers chosen from the whole numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *with repeats allowed*.
 - (a) Give a list of two numbers with the largest sample variance such a list can possibly have.
 - (b) Give a list of two numbers with the smallest sample variance such a list can possibly have.
 - (c) Does either part (a) or (b) have more than one correct answer?
 4. An experiment to study the lifetime (in hours) for a certain type of component involved putting ten components into operation and observing them for 100 hours. Eight of the ten components failed during that period, and those lifetimes were recorded. Denote

the lifetimes of the two components still functioning after 100 hours by 100+. The resulting sample observations were:

48, 79, 100+, 35, 92, 86, 57, 100+, 17, 29

Which of the measures of location/center can be calculated, and what are the values of those measurements?

(Note: The two values 100+ in this data set are said to be “right-censored”; this means that we do not know the exact lifetime of these components, only that the components must fail sometime after 100 hours (i.e., fail to the right of 100 hours). Data of this type are very common with life studies in medical trials as well as in statistical reliability, where engineers test the lifetimes of high quality materials that do not easily fail.)

5. The following frequency distribution of storm duration (in minutes) for 74 storms appeared in the article “Lightning Phenomenology in the Tampa Bay Area” (*J. of Geophysical Research* (1984): 789-805).

Storm duration in minutes, x	frequency	Storm duration in minutes, x	frequency
$0 \leq x < 25$	1	$150 \leq x < 175$	5
$25 \leq x < 50$	17	$175 \leq x < 200$	4
$50 \leq x < 75$	14	$200 \leq x < 225$	3
$75 \leq x < 100$	11	$225 \leq x < 250$	2
$100 \leq x < 125$	8	$250 \leq x < 275$	0
$125 \leq x < 150$	8	$275 \leq x < 300$	1

- (a) Sketch a histogram from the table above.
- (b) As a measure of center for the distribution, can you identify the interval in the above table that contains the median?
- (c) From the information provided in the table, can you compute the mean? Explain.