

Class 11: Topic 10: More Summary Measures and Graphs

Held: Friday, 15 February 2008

Summary: We consider more ways to summarize numeric data.

Notes:

- I've received a number of "I'm sick and won't be in class" notices. Please take care of yourselves.
- When you're sick, I would like you to try to arrange with a friend to bring homework to class. I know that's not always possible, but do the best you can.
- The exam is next Wednesday in both 3819 and 3820.
- There is no homework for Monday. Go over the sample exam and the chapters and be prepared to ask questions.
- You may find yourself a bit suspicious of the data that we're using for activity 10-3, but deal with it.
- Preparation: Overview and Preliminaries for Topic 10.
- Handouts: R notes on topic 10 and Sample Exam.
- Due: Activities 9-14, 9-15, 9-20, 9-23.

Overview:

- Some past topics: Spread, IQR, etc.
- Visualizing summary statistics.

Notes on Spread

We've seen at least three measures of the spread of a distribution.

- What are they?
- Why might one use one measure rather than another?

Notes on Inter-quartile Range

Some of you observed that the book and R give different answers for some IQR computations. Note that IQR can be computed in slightly different ways, which can have an effect on the result. (And different stats packages take different approaches.)

Let's use the distribution 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9 as our example.

Officially, the lower quartile is a number such that $1/4$ of the values are smaller and $1/4$ of the values are larger. (It's not always possible to find such a value; for example, if all the values are the same, you can't divide this evenly.) Similarly, the upper quartile is a number such that $3/4$ of the values are smaller and $1/4$ of the values are larger.

We can start by finding those values directly.

- Since there are 18 numbers in this distribution, it's impossible to find a single number such that exactly $1/4$ of the values are smaller and exactly $3/4$ are larger.
- Arguably, the fifth value in the sequence will be close enough for the lower quartile, since there should be 4 smaller values and 13 larger values.
- Similarly, the 14th value would be appropriate for the upper quartile.
- Using that strategy, we get an IQR of $7-3$ or 4 .

Our book tells us a different strategy for computing the IQR. First, compute the median and then compute the median of each half. However, our book is vague on what you do when the median is repeated.

- Suppose we just split the 18 values in the middle.
 - The left half is 1,1,2,2,3,3,4,4,5
 - The right half is 5,6,6,7,7,8,8,9,9
- Conveniently, there are nine values in each half.
 - The middle of the left half is 3
 - The middle of the right half is 7
- Hence, the IQR is 4

So, how did our book end up with 4.5 as the answer? Here's my guess. We did an odd thing in the analysis above: We put one 5 on each half. Arguably, both should go on the same side.

- Suppose we put the 5's in the lower half
 - The left half is 1,1,2,2,3,3,4,4,5,5
 - The right half is 6,6,7,7,8,8,9,9
- Now, there are ten values in the lower half and eight in the upper half. (Yeah, I don't like it either, but it's a technique that many packages still use.)
- The median of the lower half is still 3, since both the 5th and 6th values in that half are 3.
- The median value of the upper half is now 7.5, since it falls between 7 and 8.
- Hence, the IQR is 4.5

Visualizing Summary Statistics

- The five-number summary (min, lower quartile, median, upper quartile, max) is a common way of looking at data.
- But it's very "textual"
- Sometimes we find it easier to consider these same summary statistics visually.
- A boxplot (or box-and-whiskers plot) is one common way of looking at numeric data.

Copyright © 2008 Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.