

## Section 3.4: Five-Number Summary

### 1 The Five Number Summary

**Definition 1** For any set of data the *five-number summary* is, in order, the five summary statistics:

*minimum,  $Q_1$ , median,  $Q_3$ , maximum.*

**Example 2** Find the five-number summary for the Chicago Bull's salaries.

Obs	Player	Salary
1	Michael Jordan	\$33,140,000
2	Ron Harper	\$4,560,000
3	Toni Kukoc	\$4,560,000
4	Dennis Rodman	\$4,500,000
5	Luc Longley	\$3,184,900
6	Scottie Pippen	\$2,775,000
7	Bill Wennington	\$1,800,000
8	Scott Burrell	\$1,430,000
9	Randy Brown	\$1,260,000
10	Robert Parish	\$1,150,000
11	Jason Caffey	\$850,920
12	Steve Kerr	\$750,000
13	Keith Booth	\$597,600
14	Jud Buechler	\$500,000
15	Joe Kleine	\$272,250

Chicago Bulls Salaries 1997-1998 Season

The five number summary for the player salaries for the 1997-1998 Chicago Bulls is

	Excel	TI-83/84
minimum	\$272,250	\$272,250
$Q_1$	\$800,460	\$750,000
median	\$1,430,000	\$1,430,000
$Q_3$	\$3,842,450	\$4,500,000
maximum	\$33,140,000	\$33,140,000

Analysis Variable : Salary Salary								
Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean	Std Dev	10th Pctl	80th Pctl
272250.00	750000.00	1430000.00	4500000.00	33140000.00	4088711.33	8182474.38	500000.00	4530000.00

SAS output

## 2 IQR

The range is very susceptible to unusually large or small values in a data set. A single extreme value skews the range of a set of data. The interquartile range (IQR) is much more resistant to skew. The IQR measures the range of the central 50% of the data.

**Definition 3** For a given set of data, the **IQR** is the (positive or occasionally 0) difference between  $Q_3$  and  $Q_1$  in a quantitative data set.

**Example 4** For the player salaries of the 1997-1998 Chicago Bulls the range is  $33140000 - 272250 = 32867750$  and the  $IQR = 3842450 - 800460 = 3041990$ .

## 3 Outliers and IQR

We consider values that are unusually large or small compared to the rest of a data set to be **outliers**. We formally define unusually large or small via the IQR.

$$\text{lower outlier boundary} = Q_1 - 1.5 * IQR$$

$$\text{upper outlier boundary} = Q_3 + 1.5 * IQR$$

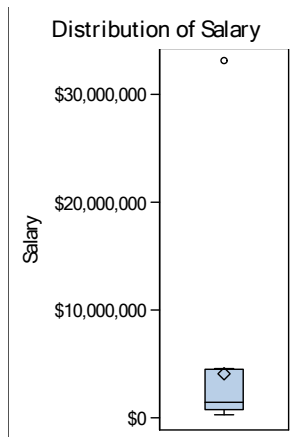
Values that fall outside these boundaries are outliers.

**Problem 5** Determine the salary outliers for the '97-'98 Bulls roster.

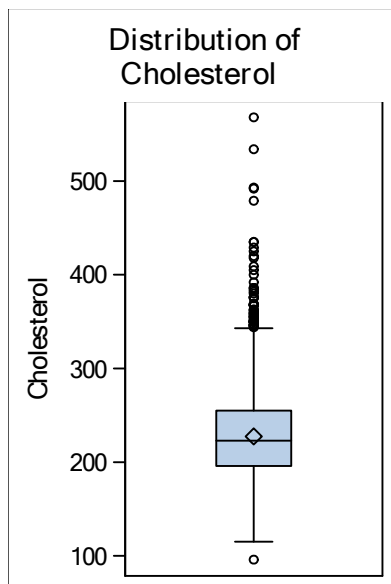
We've already determined that  $Q_1 = \$800,460$ ,  $Q_3 = \$3,842,450$  and  $IQR = 3842450 - 800460 = 3041990$ . So, the lower outlier boundary is  $800460 - 1.5 * 3041990 = -3762500$ . Clearly, nothing falls below this lower outlier boundary. The upper outlier boundary is  $3842450 + 1.5 * 3041990 = 8405400$ . Only Jordan's salary is larger than \$8,405,400. The only outlier in this data set is the salary for Michael Jordan.

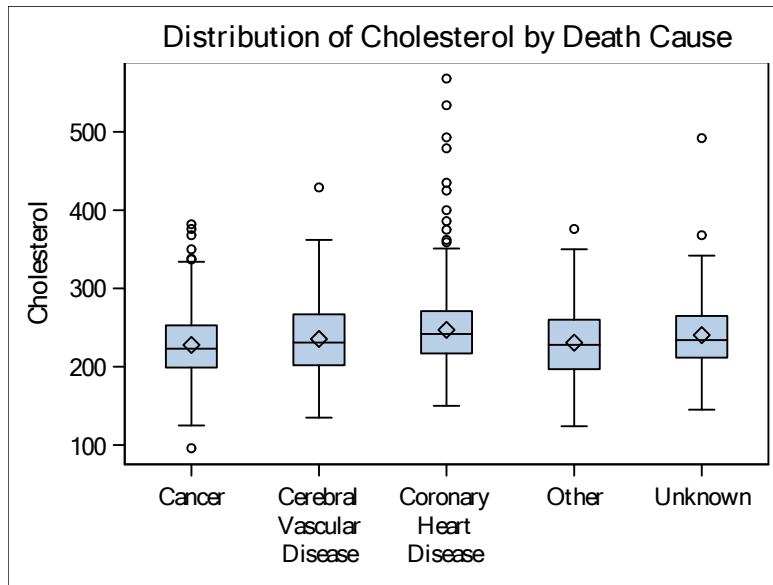
## 4 Box and Whisker Plots

A modified box and whisker plot is a graphical representation of  $Q_1$ ,  $Q_3$ , the mean, the median and outlier boundaries. The bottom and top of the box represent  $Q_1$  and  $Q_3$ . The line inside the box is the median value. The mark inside the box represents the mean. The whiskers represent the extreme values that are not outliers. Points outside these whiskers are outliers.



The box and whisker plot for the '97-'98 Chicago Bull's salaries





The modified box and whisker plot illustrates outliers. Another type of box plot is the standard box plot. The whiskers in the standard box plot represent the min and max values in the data set. Thus, no outliers are represented. My preference is for the modified box and whisker plot (which are the examples given above).

## 5 Exercises

1. Kokoska 3rd edition Section 3.4: 3.110a, 3.114a, b,