Problem 1 – Yankees VS. Mets

Visually comparing distributions

Below are two data sets that represent the 2008 salaries for the New York Yankees and the New York Mets, the two baseball teams with the highest salary totals.

New York Yankees   New York Mets

Player	Salary	Player	Salary
Alex Rodriguez	$28,000,000	Carlos Beltran	$18,622,809
Jason Giambi	$23,428,571	Johan Santana	$16,984,216
Derek Jeter	$21,600,000	Carlos Delgado	$16,000,000
Andy Pettitte	$16,000,000	Pedro Martinez	$11,813,351
Bobby Abreu	$16,000,000	Billy Wagner	$10,500,000
Richie Sexson	$15,500,000	Moises Alou	$7,500,000
Mariano Rivera	$15,000,000	Orlando Hernandez	$7,000,000
Jorge Posada	$13,100,000	Oliver Perez	$6,500,000
Johnny Damon	$13,000,000	Luis Castillo	$6,250,000
Hideki Matsui	$13,000,000	David Wright	$5,250,000
Ivan Rodriguez	$12,379,883	Brian Schneider	$4,900,000
Mike Mussina	$11,071,029	Jose Reyes	$4,375,000
Carl Pavano	$11,000,000	Scott Schoeneweis	$3,600,000
Chien-Ming Wang	$4,000,000	Ryan Church	$2,000,000
Xavier Nady	$3,350,000	Mike Pelfrey	$1,987,500
Robinson Cano	$3,000,000	Ramon Castro	$1,975,000
Damaso Marte	$2,150,000	Endy Chavez	$1,800,000
Jose Molina	$1,875,000	Matt Wise	$1,200,000
Wilson Betemit	$1,165,000	Aaron Heilman	$1,200,000
Brian Bruney	$725,000	Marlon Anderson	$1,050,000
Billy Traber	$500,000	Pedro Feliciano	$1,025,000
Melky Cabrera	$461,200	Damion Easley	$950,000
Jonathan Albaladejo	$393,225	Duaner Sanchez	$850,000
Humberto Sanchez	$390,000	John Maine	$450,000
Joba Chamberlain	$390,000	Ambiorix Burgos	$415,000
		Angel Pagan	$401,500
		Joe Smith	$398,000
		Jason Vargas	$394,000

- Without graphing, describe each data set. Think about the center and spread.

Comparing distributions using histograms

Now, on your handheld open the document *YankeesMets*. The first problem has the two lists of data from the previous page entered. On page 1.4, graph each data set, **ysalary** and **metsalary**, as a histogram.

> The graphs have been created, BUT they do not have the same scale or the same values on the horizontal axis. These needs to be adjusted so that the two data sets can be compared.
>
> Press **MENU > Plot Properties > Histogram properties > Bin settings**. Change the Width to 2 and the Alignment to 0. This has to be done for each histogram.
>
> The window also needs to be adjusted so the horizontal and vertical axes are identical. Press **MENU > Window/Zoom > Window Settings**. Change XMin to 0, XMax to 32 and YMax to 15.

- The two histograms can now be compared. Include the shape, spread, mean, and median in your comparison.

- Why is the same scale necessary?

- What would the graphs look like if the YMax value is changed – either increased or decreased? If the width of the bars is changed?

- Does changing the bars or the YMax value affect the comparison?

- How should one determine the width of the bar of a histogram?

Comparing distributions using box plots

Sometimes, a box plot is a more appropriate or useful graph to use to compare two data sets. On page 1.7, graph each data set, **ysalary** and **metsalary**, as a box plot.

> Once again, the graphs have been created, BUT they do not have the same values on the horizontal axis.
>
> The window needs to be adjusted so the horizontal axes are identical. Press **MENU > Window/Zoom > Window Settings**. Change XMin to 0 and XMax to 29.

- What do the "dots" on the Met salary graph indicate? Why are there none for the Yankees?

- Notice that with both data sets the gap between the third player and the fourth player is about 5 million dollars. Why does this create outliers for the Mets and not for the Yankees?

- What can be said about the bottom 25% of both teams?

- Compare the two box plots. Include shape, spread, and the five-number summary in your comparison.

- If you were an average player, which team would you like to play for based on the box plot? Why?

Part 5 – Drawing conclusions

What conclusions can you draw about the salaries of the baseball players on the Yankees and the Mets? Each conclusion needs to be supported by a number or a graph.

Problem 2 – East VS. West

Problem 2 on your handheld has data about teenage birthrates for the 50 states and District of Columbia. It is split into Eastern and Western states. You are the Health Commissioner for the United States. Analyze the data sets and compare the two distributions. What conclusions can you make about the birthrates for the Eastern and Western states based on your findings? Use graphs and numerical summaries to support your conclusions.