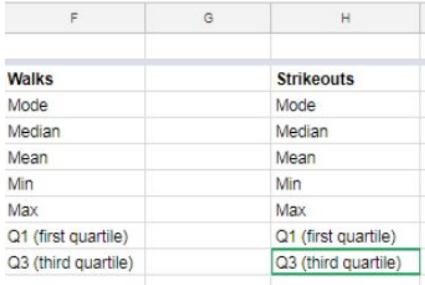


Name \_\_\_\_\_

## Probability and Statistics

### Baseball Data Google Sheets Instructions

1. Make a copy of the Baseball Data Shared with you in Google Classroom or your WRemail.
2. Under View -> Freeze and 1 Row (this will lock your titles in row 1)
3. Start in cell F2 write in Walks (bold print) and continue the rest of the words as follows, using bold print also on Strikeouts. You may have to make the columns a bit larger to fit the names for the quartiles. Do this by going all the way up to the line that separates the F and G columns and when the little line goes bold and cursor becomes a two way arrow, slide the arrow over to make the column larger. Do the same for column H.  


F	G	H
Walks	Strikeouts	
Mode	Mode	
Median	Median	
Mean	Mean	
Min	Min	
Max	Max	
Q1 (first quartile)	Q1 (first quartile)	
Q3 (third quartile)	Q3 (third quartile)	

4. In G3, type in =MODE(A2:A143) or instead of typing A2:A143, you can just highlight those cells and they will automatically be put in. You can do this for all of the following, but I will not repeat it.
5. In G4, type =MEDIAN(A2:A143)
6. In G5, type =AVERAGE(A2:A143) To change the decimals to only 2 places, highlight the cell then towards the top hit the .0 with left arrow underneath till you only have 2 decimal places left.
7. In G6, type =MIN(A2:A143)
8. In G7, type =MAX(A2:A143)
9. In G8, type =QUARTILE(A2:A143,1)
10. In G9, type =QUARTILE(A2:A143,3)
11. Now let's say we forgot to do the standard deviation and want to put it in right after mean, highlight cells F6, G6 and H6. Choose Insert -> cells and shift down. Now type in Standard deviation in cells F6 and H6. In cell G6 type =STDEV(A2:A143) change to only 2 decimals
12. Now do the same in column I for the strikeouts (so use data in B2:B143) I just copied and pasted from column G into column I and changed all the Cs (that's what my computer changed them to for the data) to B's. Up to you on how you do it though.
13. Now let's use the Statistics Add-on to make the box-in-whisker plot (Boxplot). I want you to make ONLY a boxplot of Walks and Strikeouts. Go to the statistics add on and choose describe Data. Add the variables Walks and Strikeouts and under Continuous Variable Options only check mark the boxplot, then Create.
14. Double check your info with the newly created sheet.

15. Rename the univariate1 sheet to “boxplots” by hitting the down arrow on the sheet on the bottom of the page and choose Rename. Rename your original sheet1 to “data”

16. Using the Describe Data again create a sheet with histograms as the only charts and rename that sheet as “histograms”.

17. Click on the Walks histogram and I want you to change the bars colors. In the top right hand corner of the chart there are three vertical dots, hit that and choose to edit chart...

Go to customize and click the  $\vee$  (drop down menu) for Series, go to color and choose a different color.

18. Under “color” is Type, and choose “value” from the drop down menu. This lets you know the frequency of the “bins” or bars.

19. Under “chart and axis titles, put a title of Walks on your chart. You may change font and color of text if you’d like.

20. Do the steps 17-19 for Strikeouts, but choose a different color for bars.

21. To make a scatterplot for Strikeouts vs ERA, highlight strikeout column including the words strikeouts in the first row. Then hold CTRL while highlighting the ERA column and again including the words ERA (by highlighting the words, it’ll help with labels).

22. While the two columns are highlighted, go to insert (on top of page) and chart. It’ll probably pop up a different chart. Move it to start the top left corner at cell F17.

23. Change the chart type to scatterplots by hitting the drop down menu in the chart editor and choosing the first plot under scatter.

24. On the bottom of the chart editor hit “Use the old chart editor”.

25. Make a title, type “Strikeouts” for horizontal axis title, you may change colors and shape of dots on scatterplot under series, and on the very bottom of the chart editor under trendline choose linear, choose “use equation” for Label and show  $R^2$ . Hit “update” when finished.

26. Starting in Cell F12 write Strikeouts vs ERA in bold and fill in the following.

Strikeouts vs ERA
Slope
Intercepts
Correlation

27. In G13 type =SLOPE(D2:D143,B2:B143). This finds the slope of the ERA (y) and Strikeouts (x)

28. In G14 type =INTERCEPT(D2:D143,B2:B143).

29. In G15 type =CORREL(D2:D143,B2:B143) remember this is your R (correlation coefficient)

30. Check the equation you got from step 25 and your slope and intercept you got from 27 and

28. Also if you take the square root of the  $R^2$  from step 25 you should get your correlation from

29. (remember since it has a negative correlation, the R will be negative while the  $R^2$  will be positive)