Descriptive Statistics Project
Is there a home field advantage in major league baseball?

DUE at start of class *(first 5 minutes of class)* **on date posted on website**
There may be other due dates that week also – start early & plan your work accordingly.

Are baseball teams more likely to win when they play on their home field than when they play away from home on their opponent's field? We will study the most recent complete regular season (2018) data for all Major League baseball teams.

Objective: To graph data and calculate statistics for the number of home games won by each team (Part A) and the number of away games won by each team (Part B) and analyze the results to determine if the data show evidence of a home field advantage (Part C).

This is a collaborative project. Students must work in a group (2 to 4 people to a group.) Each person in the group must contribute to the worksheet calculations in part A or B, discuss and compare results, and contribute to the interpretation of graphical and numerical results in part C. *Students absent on the day project work is started in groups in class need to see the instructor when they return to class; any absent student will do the project individually, unless the instructor can find another student with such absence to partner together.*

Instructions

Hand in ONE copy of the project for the group.
(in order, stapled, with all questions answered)

Checklist: It should include ALL of the following:

- **Cover Page with the project title and the names of all people in the group.**
  If work was divided up, explain who worked on which parts.
  Cover sheet is used for grading and comments. Missing cover sheet loses 2 points for not following instructions.
- **Parts A & B: descriptive statistics worksheets**
  Hand in the project worksheets with the calculations and graphs.
- **Part C: Typed Analysis & Conclusions**
  Use the graphs and descriptive statistics to draw conclusions and justify them, to answer questions in part C. Only Part C must be typed.

All group members should have a copy as backup. If only one group member has a copy and that person drops the class or misses the deadline, it adversely affects the grade of the whole group.

TEAM	HOME WINS	AWAY WINS
Boston	57	51
NY Yankees	53	47
Tampa Bay	51	39
Toronto	40	33
Baltimore	28	19
Cleveland	49	42
Minnesota	49	29
Detroit	38	26
Chi White Sox	30	32
Kansas City	32	26
Houston	46	57
Oakland	50	47
Seattle	45	44
LA Angels	42	38
Texas	34	33
Atlanta	43	47
Washington	41	41
Philadelphia	49	31
NY Mets	37	40
Miami	38	25
Milwaukee	51	44
Chi Cubs	51	44
St. Louis	43	45
Pittsburgh	44	38
Cincinnati	37	30
Coloradox	47	44
LA Dodgers	44	47
Arizona	40	42
San Francisco	**42**	**31**
San Diego	31	35

Explanation of data: The data count the number of games won when playing on the home field and the number of games won when playing away on the opponent's field.
Example: The S.F.Giants won 42 games when playing at home at their own (home) field.
The S.F.Giants won 31 games when playing away on the other team's field.
Data are counts of games won. Data are ***not*** *"scores" of points earned for any particular games.*

Parts A&B Worksheet 1 - Descriptive Statistics – Home Team Advantage Project

Find summary statistics for the data using your graphing calculator.
Consider the data as a population of all games from this season to select the appropriate standard deviation. *Round all values to the nearest tenth (1 decimal place – round accurately!)*

A1. Home Win Data	**B1. Away Win Data**
Mean _____	Mean _____
Standard Deviation _____	Standard Deviation _____
Min _____	Min _____
Q1 _____	Q1 _____
Median _____	Median _____
Q3 _____	Q3 _____
Max _____	Max _____
IQR_____	IQR_____

Create a stem and leaf plot of the data. *Stems are printed below – fill in the leaves – do not put commas between the values of the leaves - space leaves to accurately show shape of data.*

A2. Home Win Data		**B2. Away Win Data**	
1	_____	**1**	_____
2	_____	**2**	_____
3	_____	**3**	_____
4	_____	**4**	_____
5	_____	**5**	_____
6	_____	**6**	_____
7	_____	**7**	_____
8	_____	**8**	_____

A3 & B3. Draw boxplots for each set of data, neatly below; draw one above the other for easy comparison
Label each to indicate which is home data and which is away data.
Use a RULER to draw the graph TO SCALE with STRAIGHT LINES

20 25 30 35 40 45 50 55 60 games won

Part A&B Worksheet 2: Descriptive Statistics – Home Team Advantage Project

A4 & B4. For each set of a data, create a frequency histogram.

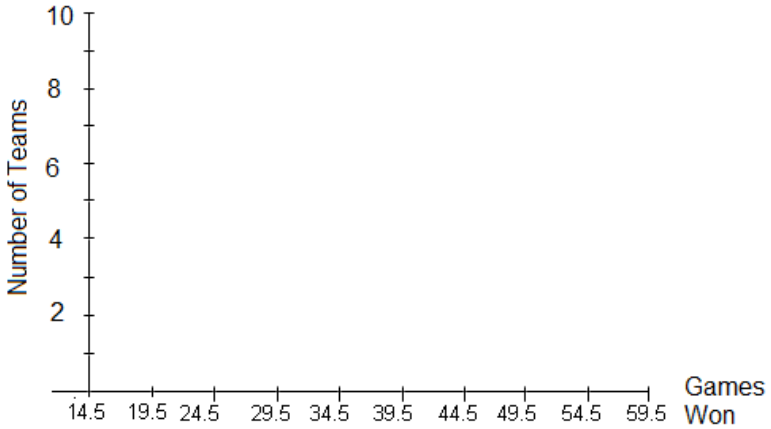
Start by completing the GROUPED frequency table using intervals below, with interval boundaries
14.5-19.5,19.5–24.5, 24.5–29.5,29.5–34.5,34.5–39.5,39.5–44.5,44.5–49.5,49.5–54.5,54.5–59.5

Neatly draw the histogram. Use a RULER to draw the graph TO SCALE with STRAIGHT LINES.

Messy or inaccurate graphs will lose points because they don't convey information clearly.
Neighboring bars should be touching; no gaps between bars unless there is no data for some interval.)

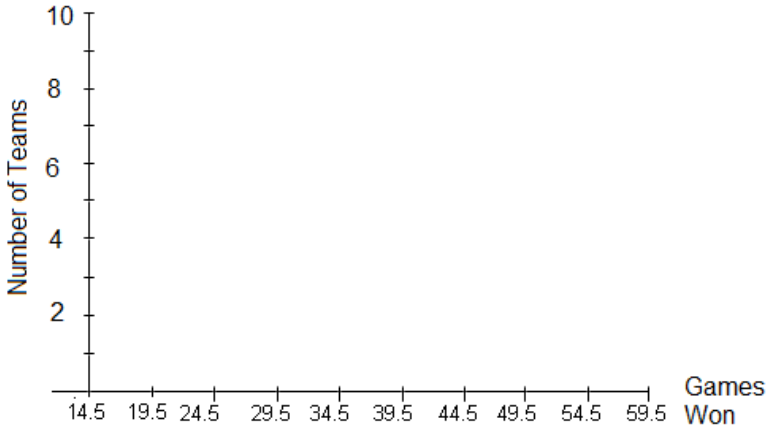
A4. HOME DATA

Games Won at HOME	Frequency



B4. AWAY DATA

Games Won AWAY	Frequency

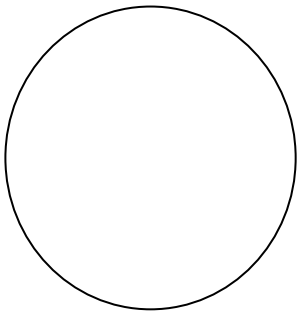


A5. & B5. Complete the table using data for all teams.
Use table to make a pie chart.

Add up the total number of games won at home for all teams and then add up total number of games won away for all teams. Find the percentages.

Pie chart should contain two reasonably accurately sized wedges for the percent of all games won at home and won away. Label pie chart to identify home vs away.

	Total Number of games won for all teams	Percent of games won
AT HOME		
AWAY		
TOTAL (home + away)		



A6.& B6. Complete the table.

How many teams won more games at home than away?

How many teams won more games away than at home?

Fill in table using counts and percent**s.**

Note: There are 20 teams in total.

	Number of Teams	Percent of Teams
Won more games at home than away		
Won more games away than at home		

Part A Worksheet 3 - Descriptive Statistics – Home Team Advantage Project

A7. Fill in the table showing sorted individual data values, frequency, relative frequency and cumulative relative frequency.

Do NOT group data into intervals. Each data value gets its own line in the table. If a value appears more than once in the data, it still has only one line in the table. The frequency column shows how many times each value appears in the data.
If a value (such as 31 or 50) does not appear in the data, it does not get a line in the table.
Round all values to 3 decimal places. Round carefully. (Hint: 0.66666666 rounds to 0.667)

Number of Games Won AT HOME	Frequency	Relative Frequency	Cumulative Relative Frequency

Use the table to find the value of the 65th percentile AND write the sentence that interprets the 65th percentile in the context of the situation.

The 65th percentile is _____ games won when playing at home.

Write the interpretation of the 65th percentile in the context of this situation:

_____

_____

_____

A8. Are there any outliers in the data for games won at home?

Show the test for outliers that uses the IQR and clearly state your conclusion based on your work.

Part B Worksheet 4 - Descriptive Statistics – Home Team Advantage Project

B7. Fpr the games won AWAY, fill in the table showing sorted individual data values, frequency, relative frequency and cumulative relative frequency.

Do NOT group data into intervals. Each data value gets its own line in the table. If a value appears more than once in the data, it still has only one line in the table. The frequency column shows how many times each value appears in the data.
If a value (such as 31 or 50) does not appear in the data, it does not get a line in the table.
Round all values to 3 decimal places. Round carefully. (Hint: 0.66666666 rounds to 0.667)

Number of Games Won AWAY	Frequency	Relative Frequency	Cumulative Relative Frequency

Use the table to find the value of the 65th percentile AND write the sentence that interprets the 65th percentile in the context of the situation.

The 65th percentile is _____ games won when playing away.

Write the interpretation of the 65th percentile in the context of this situation:

_____

_____

_____

B8. Are there any outliers in the data for games won away?

Show the test for outliers that uses the IQR and clearly state your conclusion based on your work.

Part C: SUMMARY OF RESULTS

Based on the evidence in the data, do the data support or not support the idea that there is a home team advantage in Major League Baseball?

The written summaries in this section need to be TYPED, not handwritten.

If you do not have access to a computer and printer at home, there are places on campus that provide access. Use complete sentences. Proofread your analysis so that your writing is grammatically correct and makes sense. If you need help editing your writing (not the math) you can get assistance at the Reading and Writing Center in ATC309 https://www.deanza.edu/studentsuccess/wrc/

C1. <u>Based on all the graphs</u>, does there appear to be a home field advantage?

Compare the graphs (box plots, histograms, and pie chart)
Explain in complete sentences, specifically citing what you see in *each* of the graphs to support your conclusion.

> *C1 is the hardest part to write for some students. Just saying "the graph shows that teams won more games at home "is not enough detail or information. Explain how the graph visually shows what you are claiming. (Think about how you would explain what each graph shows to somebody who did not understand what the graphs shows about the data.)*

C2. <u>Based on all the summary statistics</u>, does there appear to be a home field advantage?

Compare the numerical summary statistics that measure <u>location</u> of the data
(means, medians, quartiles, min, max., 65th percentiles, and % of games won at home vs away)
Explain in complete sentences, specifically citing what you see in the numerical summary statistics to support your conclusion.

NOTE:

You will be graded on the quality of your answers to the analysis in Part C. Be thoughtful and careful writing up your conclusions. In order to earn all the points for the written analysis, it must be complete and accurate with sufficient detail to justify your conclusions based on the evidence in the data. A poor analysis can lose points even if the technical work is correct.

If you need help editing your written work to make grammatical sense with quality writing, you can visit the Reading and Writing Center for assistance: <u>https://www.deanza.edu/studentsuccess/wrc/</u>

This table summarizes the points allocated for each part of this project:

A1 & B1 (total)	A2 & B2 (total)	A3 & B3 (total)	A4 & B4 (total)	A5 & B5 (total)	A6 & B6 (total)	A7	B7	A8	B8	C1	C2	Total Points
Summary Statistics	Stem & Leaf	Box Plots	Tables & Histograms	Table & Pie Chart	Table	Table & %ile & Interpret		Outliers		Written Analysis		
5	4	4	6	4	3	4	4	3	3	5	5	50