Practice Project Overview

Estimated time needed: 1 hour

Introduction

This hands-on project focuses on refreshing your Microsoft Excel data analysis skills and applying them to a real-life scenario. You will be calculating the measures of central tendency and dispersion, examining the probability distribution, grouping data using pivot tables to gather insights, and finally performing a regression analysis and interpreting it.

This reading gives you an overview of the scenario and tasks to be performed. The project instructions will provide you with a data set to work on and the steps to perform each task.

Assignment Scenario

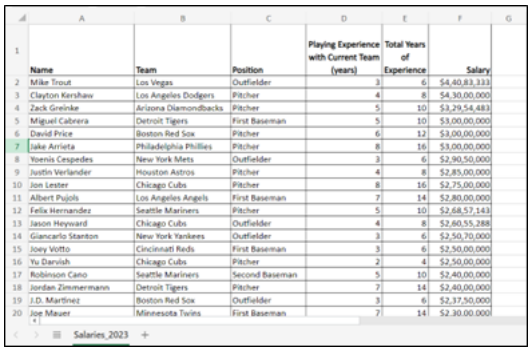
The National League Baseball (NLB) season of 2023 saw a diverse range of talented players across various positions, with salaries reflecting their contributions. The top-paid players demonstrated their exceptional skills, while younger players continued to develop and showcase their talent, shaping the future of the league.

The salaries also vary depending on the team, and position in the team. You have been hired by the league as their Data Analyst. You will be presenting statistical reports that will help the league or team management make the decisions to project their remunerations for the new contract.

In this practice, you will:

- Calculate the measures of central tendency: Mean, median, and mode.
- Calculate the measures of dispersion: Standard deviation and variance.
- Find the minimum, maximum, percentiles, and quartiles.
- Create a histogram.
- Create a pivot table.
- Perform a multiple linear regression and interpret the outcome.

You will be provided with a data set showing the salaries of 877 players, their team names, their positions, and their experience. The screenshot of the data set is displayed below.

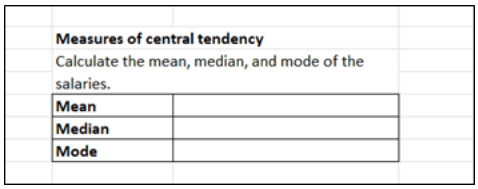


	A	B	C	D	E	F	G
1	Name	Team	Position	Playing Experience with Current Team (years)	Total Years of Experience	Salary	
2	Mike Trout	Los Vegas	Outfielder	3	6	$4,40,83,333	
3	Clayton Kershaw	Los Angeles Dodgers	Pitcher	4	8	$4,30,00,000	
4	Zack Greinke	Arizona Diamondbacks	Pitcher	5	10	$3,29,54,483	
5	Miguel Cabrera	Detroit Tigers	First Baseman	5	10	$3,00,00,000	
6	David Price	Boston Red Sox	Pitcher	6	12	$3,00,00,000	
7	Jake Arrieta	Philadelphia Phillies	Pitcher	8	16	$3,00,00,000	
8	Yoenis Cespedes	New York Mets	Outfielder	3	6	$2,90,50,000	
9	Justin Verlander	Houston Astros	Pitcher	4	8	$2,85,00,000	
10	Jon Lester	Chicago Cubs	Pitcher	8	16	$2,75,00,000	
11	Albert Pujols	Los Angeles Angels	First Baseman	7	14	$2,80,00,000	
12	Felix Hernandez	Seattle Mariners	Pitcher	5	10	$2,68,57,143	
13	Jason Heyward	Chicago Cubs	Outfielder	4	8	$2,60,55,288	
14	Giancarlo Stanton	New York Yankees	Outfielder	3	6	$2,50,70,000	
15	Joey Votto	Cincinnati Reds	First Baseman	3	6	$2,50,00,000	
16	Yu Darvish	Chicago Cubs	Pitcher	2	4	$2,50,00,000	
17	Robinson Cano	Seattle Mariners	Second Baseman	5	10	$2,40,00,000	
18	Jordan Zimmermann	Detroit Tigers	Pitcher	7	14	$2,40,00,000	
19	J.D. Martinez	Boston Red Sox	Outfielder	3	6	$2,37,50,000	
20	Joe Mauer	Minnesota Twins	First Baseman	7	14	$2,30,00,000	

Salaries_2023 +

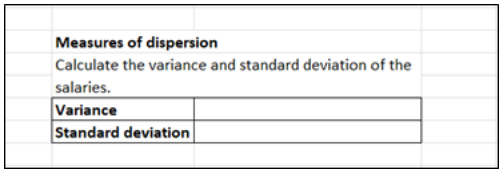
Tasks to perform:

1. **Measures of central tendency:** Calculate the mean, median, and mode of the salaries. Perform the calculations in the cells provided.



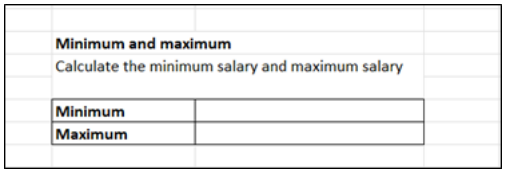
Measures of central tendency	
Calculate the mean, median, and mode of the salaries.	
Mean	
Median	
Mode	

2. **Measures of dispersion:** Calculate the variance and standard deviation for salaries. Perform the calculations in the cells provided.



Measures of dispersion	
Calculate the variance and standard deviation of the salaries.	
Variance	
Standard deviation	

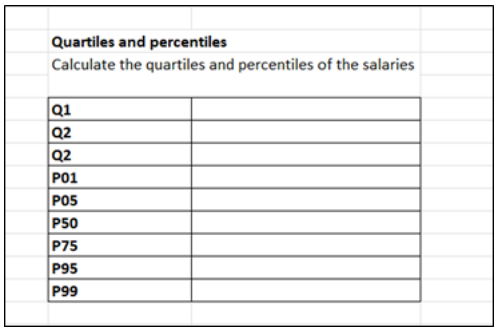
3. **Minimum and maximum:** Find the minimum and maximum salary. Perform the calculations in the cells provided.



Minimum and maximum	
Calculate the minimum salary and maximum salary	
Minimum	
Maximum	

4. **Quartiles and percentiles:** Calculate the following quartiles and percentiles in the cells provided for the calculation:

- First, second, and third quartiles of the salaries (Q1, Q2, Q3)
- 1, 5, 50, 75, 95, and 99 percentiles of the salaries (P01, P05, P50, P75, P95, P99)



Quartiles and percentiles

Calculate the quartiles and percentiles of the salaries

Q1	
Q2	
Q2	
P01	
P05	
P50	
P75	
P95	
P99	

5. **Histogram:** Create a histogram showing the probability distribution of salaries. What does it tell you about the distribution of data?

6. **Pivot charts:** Create two pivot charts in two separate worksheets to obtain the following information.
   a. Find the team paying the highest salaries to its players.
   b. Find the position being paid the highest salary.

7. **Multiple linear regression analysis:**
   a) Perform a multiple linear regression analysis to predict the salaries for the next year. The salaries depend on the playing experience with the current team and the total years of experience.
   b) Interpret the output. Record the R-squared value and Significance F of the model and the p-values for each independent variable and explain what these values signify.

The detailed step list is provided in the Practice Project Guide.

Author(s)

- Dr. Rajendra Patil

Other Contributors

- Sangeeta Srinivasan
- Rashi Kapoor