

Basic Statistical Measures for Sports Fans

Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Completion Percentage	60.2	55.1	64.3	67.8	59.2	70.1	58.9	61.7	68.2

Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Rushing Yards	304	506	123	453	395	402	491	218	286

Mean?

Median?

Mode?

Range?

Basic Statistical Measures for Sports Fans-Introduction Part 1

The field of statistics is a collection of methods used to organize, display, describe, and analyze a set of data (commonly referred to as a data set). Four main numbers are used to present a statistical representation of a data set. These are:

- Mean-The mathematical average of all terms in the data set. To calculate the mean, add up the values of all the terms in the data set and then divide by the number of terms.
- Median-The middle of all the terms in the data set. To find the median, list the terms in the data set in numerical order and identify which term appears in the middle of the list. If there are an even number of terms in the data set, then the median is the average of the two terms in the middle.
- Mode-The most common number in the data set. To find the mode, list the terms in the data set in numerical order and identify which in the data set occurs most often. A data set can have multiple modes. A data set with one mode is called unimodal, one with two modes is called bimodal, one with three modes is called trimodal and one with four or more modes is called multi modal. If a data set has does not have any repeating terms, it has no modes.
- Range-The difference between the maximum term and the minimum term in a data set. To calculate the range, list the terms in the data set in numerical order and subtract the smallest term from the largest term.

The following page shows the calculations for each of these measures based on the data set below:

4	2	7	5	9	7	7	5	6
---	---	---	---	---	---	---	---	---

Basic Statistical Measures for Sports Fans-Introduction Part 2

To find the four statistical measures for the data set shown on the previous page:

- List the terms in the data set in numerical order (this step can be skipped if only the mean is being calculated):

2	4	5	5	6	7	7	7	9
---	---	---	---	---	---	---	---	---

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 52
 - Divide this total by the number of terms in the data set which is 9
 - The mean equals $52/9$ which is 5.778
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 6 since there are 4 terms above it and 4 terms below it

Note that in this case the mean and median are close. This occurs when the data set does not contain any outliers which are terms that differ significantly from the others. If the mean and the median are exactly the same, the distribution of the data set is called symmetrical.

- To find the mode:
 - Determine which term in the data set appears the most often
 - The mode is 7 since it appears 3 times (trimodal)
- To calculate the range:
 - Determine the largest term in the data set which is 9
 - Determine the smallest term in the data set which is 2
 - Subtract 2 from 9 yielding 7 which is the range

In the world of sports, statistics are used to rate players, decide on who gets cut from the team, decide who plays on which days and against which teams and with which teammates. Statistics are also used in player salary arbitration, and in making draft picks.

Basic Statistical Measures for Sports Fans-Example

Example: A football team has a defensive front seven with the following weights:

200	380	210	220	205	200	200
-----	-----	-----	-----	-----	-----	-----

Determine the mean, median, mode (and its type), and range for this data set. If the mean and median are not close, identify the reason.

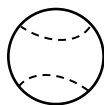
- List the terms in the data set in numerical order:

200	200	200	205	210	220	380
-----	-----	-----	-----	-----	-----	-----

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 1615
 - Divide this total by the number of terms in the data set which is 7
 - The mean equals $1615/7$ which is 230.71
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 205 (there are 3 terms above it and 3 terms below it)
- To find the mode:
 - Determine which term in the data set appears the most often
 - The mode is 200 since it appears 3 times (trimodal)
- To calculate the range:
 - Determine the largest term in the data set which is 380
 - Determine the smallest term in the data set which is 200
 - Subtract 200 from 380 yielding 180 which is the range

The mean and median are not close because the player weighing 380 pounds is a significant outlier from the other six player's weights.

Basic Statistical Measures for Sports Fans-Problem 1



Problem 1: A baseball team has a starting lineup of nine players with ages as shown below:

27	31	36	26	25	32	34	29	30
----	----	----	----	----	----	----	----	----

Determine the mean, median, mode (and its type), and range for this data set. Discuss the relationship between the mean and the median.

Basic Statistical Measures for Sports Fans-Problem 2

Problem 2: Ten players on a basketball team scored the following number of points in a game:

8	9	0	9	6	4	30	5	3	1
---	---	---	---	---	---	----	---	---	---

Determine the mean, median, mode (and type) , and range for this data set. Discuss the relationship between the mean and the median.

Basic Statistical Measures for Sports Fans-Problem 3

Problem 3: An NFL team is preparing for the college draft and is focused on Quarterbacks. The team wants to narrow its list to only those from a group of nine that exceed the median of the group in both pass completion percentage and rushing yards for the last college season. Based on the information below which Quarterbacks would be on their short list?

Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Completion Percentage	60.2	55.1	64.3	67.8	59.2	70.1	58.9	61.7	68.2

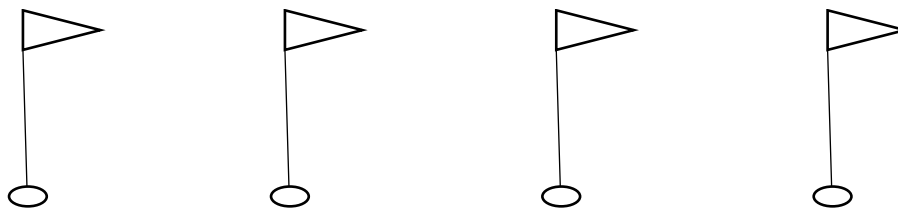
Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Rushing Yards	304	506	123	453	395	402	491	218	286

Basic Statistical Measures for Sports Fans-Problem 4

Problem 4: A major league baseball team wants to invite some minor league pitchers to spring training. The team decides it wants to select pitchers with lower Earned Run Averages (ERA). This metric is the average number of earned runs allowed in a nine inning game. Based on the data set below, which pitchers would be selected if the team invites pitchers with ERA's below the mean of the group? Which pitchers would be selected if the team invites players with ERA's below the median of the group?

Metric/Player	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
ERA	2.26	3.24	5.61	6.45	1.89	4.12	3.88	2.79	4.86	7.12	3.47

Basic Statistical Measures for Sports Fans-Problem 5



Problem 5: Two golfers are playing a match. Their scores for their last nine rounds are shown below. If the score for each player in this match is equal to their mean score for the last nine rounds which player wins? If the score for each player in this match is equal to their mode score for the last nine rounds which player wins? Note remember that in golf, low score wins.

Player 1	74	68	78	74	80	73	74	75	70
Player 2	72	75	79	81	78	82	80	77	72

Basic Statistical Measures for Sports Fans-Problem 6

Problem 6: The yearly salaries (in millions) for the starting five players on an NBA basketball team are shown below already in numerical order. If P2 gets a raise to \$2.8 million, how do you think the mean salary for the team will change? How do you think the median salary for the team will change? Show calculations to support your answer.

Metric/Player	P1	P2	P3	P4	P5
Annual Salary (\$M)	1.2	2.3	3.0	4.6	5.4

Basic Statistical Measures for Sports Fans

Answers

Basic Statistical Measures for Sports Fans-Problem 1 Answer

Problem 1: A baseball team has a starting lineup of nine players with ages as shown below:

27	31	36	26	25	32	34	29	30
----	----	----	----	----	----	----	----	----

Determine the mean, median, mode (and its type), and range for this data set. Discuss the relationship between the mean and the median.

Problem 1 Answer:

- List the terms in the data set in numerical order:

25	26	27	29	30	31	32	34	36
----	----	----	----	----	----	----	----	----

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 270
 - Divide this total by the number of terms in the data set which is 9
 - The mean equals $270/9$ which is 30
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 30 (there are 4 terms above it and 4 terms below it)
- To find the mode:
 - Determine which term in the data set appears the most often
 - There is no mode because the data set does not have any repeating terms
- To calculate the range:
 - Determine the largest term in the data set which is 36
 - Determine the smallest term in the data set which is 25
 - Subtract 25 from 36 yielding 11 which is the range

The mean and median are exactly the same and the distribution of the data set is called symmetrical. There are no significant outliers in the data set.

Basic Statistical Measures for Sports Fans-Problem 2 Answer

Problem 2: Ten players on a basketball team scored the following number of points in a game:

8	9	0	9	6	4	30	5	3	1
---	---	---	---	---	---	----	---	---	---

Determine the mean, median, mode (and type) , and range for this data set. Discuss the relationship between the mean and the median.

Problem 2 Answer:

- List the terms in the data set in numerical order:

0	1	3	4	5	6	8	9	9	30
---	---	---	---	---	---	---	---	---	----

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 75
 - Divide this total by the number of terms in the data set which is 10
 - The mean equals $75/10$ which is 7.5
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - Since these are an even number of terms in this data set, 5 and 6 are the middle terms (there are 4 terms above these two and 4 terms below these two)
 - The median is the sum of these terms divided by 2
 - The median is $(5 + 6) / 2 = \underline{5.5}$
- To find the mode:
 - Determine which term in the data set appears the most often
 - The mode is 9 since it appears 2 times (bimodal)
- To calculate the range:
 - Determine the largest term in the data set which is 30
 - Determine the smallest term in the data set which is 0
 - Subtract 0 from 30 yielding 30 which is the range

The mean and median are not close because the player who scored 30 points is a significant outlier from the other nine players.

Basic Statistical Measures for Sports Fans-Problem 3 Answer

Problem 3: An NFL team is preparing for the college draft and is focused on Quarterbacks. The team wants to narrow its list to only those from a group of nine that exceed the median of the group in both pass completion percentage and rushing yards for the last college season. Based on the information below which Quarterbacks would be on their short list?

Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Completion Percentage	60.2	55.1	64.3	67.8	59.2	70.1	58.9	61.7	68.2

Metric	QB1	QB2	QB3	QB4	QB5	QB6	QB7	QB8	QB9
Rushing Yards	304	506	123	453	395	402	491	218	286

Problem 3 Answer:

- List the terms in the data set for Completion Percentage in numerical order:

Metric	QB2	QB7	QB5	QB1	QB8	QB3	QB4	QB9	QB6
Completion Percentage	55.1	58.9	59.2	60.2	61.7	64.3	67.8	68.2	70.1

- To find the median of this data set:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 61.7 (there are 4 terms above it and 4 terms below it)
- Determine which QB's are above the median for Completion Percentage
 - QB3, QB4, QB9, QB6
- List the terms in the data set for Rushing Yards in numerical order:

Metric	QB3	QB8	QB9	QB1	QB5	QB6	QB4	QB7	QB2
Rushing Yards	123	218	286	304	395	402	453	491	506

- To find the median of this data set:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 395 (there are 4 terms above it and 4 terms below it)
- Determine which QB's are above the median for Rushing Yards:
 - QB6, QB4, QB7, QB2
- Determine which QB's are above the median for both Completion Percentage and Rushing Yards:
 - QB4 and QB6 are above the medians in both metrics so these two Quarterbacks would be on the team's short list for the draft.

Basic Statistical Measures for Sports Fans-Problem 4 Answer

Problem 4: A major league baseball team wants to invite some minor league pitchers to spring training. The team decides it wants to select pitchers with lower Earned Run Averages (ERA). This metric is the average number of earned runs allowed in a nine inning game. Based on the data set below, which pitchers would be selected if the team invites pitchers with ERA's below the mean of the group? Which pitchers would be selected if the team invites players with ERA's below the median of the group?

Metric/Player	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
ERA	2.26	3.24	5.61	6.45	1.89	4.12	3.88	2.79	4.86	7.12	3.47

Problem 4 Answer:

- List the terms in the data set for ERA in numerical order:

Metric/Player	P5	P1	P8	P2	P11	P7	P6	P9	P3	P4	P10
ERA	1.89	2.26	2.79	3.24	3.47	3.88	4.12	4.86	5.61	6.45	7.12

- To calculate the mean for this data set:
 - Add all the terms in the data set yielding a total of 45.69
 - Divide this total by the number of terms in the data set which is 11
 - The mean equals $45.69/11$ which is 4.15
- To find the median of this data set:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 3.88 (there are 5 terms above it and 5 terms below it)

So, seven pitchers (P5, P1, P8, P2, P11, P7, P6) are below the mean and would be invited to spring training but only five pitchers (P5, P1, P8, P2, and P11) are below the median and would be invited to spring training based on this criteria.

Basic Statistical Measures for Sports Fans-Problem 5 Answer

Problem 5: Two golfers are playing a match. Their scores for their last nine rounds are shown below. If the score for each player in this match is equal to their mean score for the last nine rounds which player wins? If the score for each player in this match is equal to their mode score for the last nine rounds which player wins? Note remember that in golf, low score wins.

Player 1	74	68	78	74	80	73	74	75	70
----------	----	----	----	----	----	----	----	----	----

Player 2	72	75	79	81	78	82	80	77	72
----------	----	----	----	----	----	----	----	----	----

Problem 5 Answer:

- List the terms in the data set for Player 1 in numerical order:

Player 1	68	70	73	74	74	74	75	78	80
----------	----	----	----	----	----	----	----	----	----

- To find the median for Player 1:
 - Determine which term appears in the middle of the sorted data set
 - The median is 74 (there are 4 terms above it and 4 terms below it)
- To find the mode for Player 1:
 - Determine which term in the data set appears the most often
 - The mode is 74 since it appears 3 times (trimodal)

- List the terms in the data set for Player 2 in numerical order:

Player 2	72	72	75	77	78	79	80	81	82
----------	----	----	----	----	----	----	----	----	----

- To find the median for Player 2:
 - Determine which term appears in the middle of the sorted data set
 - The median is 78 (there are 4 terms above it and 4 terms below it)
- To find the mode for Player 2:
 - Determine which term in the data set appears the most often
 - The mode is 72 since it appears 2 times (bimodal)

So, if the score for each player equals their median for the last nine rounds, Player 1 wins with a score of 74 to 78. However, if the score for each players equals their mode for the last nine rounds, Player 2 wins with a score of 72 to 74.

Basic Statistical Measures for Sports Fans-Problem 6 Answer

Problem 6: The yearly salaries (in millions) for the starting five players on an NBA basketball team are shown below already in numerical order. If P2 gets a raise to \$2.8 million, how do you think the mean salary for the team will change? How do you think the median salary for the team will change? Show calculations to support your answer.

Metric/Player	P1	P2	P3	P4	P5
Annual Salary (\$M)	1.2	2.3	3.0	4.6	5.4

Problem 6 Answer:

The mean salary should increase because one of the entries in the data set increased. The median should remain unchanged because the new salary for P2 is still below the original median. The calculations that show this to be correct are:

For the original case:

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 16.5
 - Divide this total by the number of terms in the data set which is 5
 - The mean equals $16.5/5$ which is 3.3
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 3.0 (there are 2 terms above it and 2 terms below it)

For the revised case the data set is now:

Metric/Player	P1	P2	P3	P4	P5
Annual Salary (\$M)	1.2	2.8	3.0	4.6	5.4

- To calculate the mean:
 - Add all the terms in the data set yielding a total of 17.0
 - Divide this total by the number of terms in the data set which is 5
 - The mean equals $17.0/5$ which is 3.4
- To find the median:
 - Determine which term(s) appears in the middle of the sorted data set
 - The median is 3.0 (there are 2 terms above it and 2 terms below it)