

Problem Set 2

CSCI S-5a: Programming in R

Student: 91556812

Problem 1: Baseball Statistics

The great player Willie Mays had these career statistics:

Statistics	Value
Plate appearances	12,497
At-bats	10,881
Hits	3,283
Doubles	523
Triples	140
Home Runs	660
Bases on balls	1,464
Hit by a pitch	44
Sacrifice flies	91

Part (a)

Calculate Willie Mays' batting average. Report your result using a `cat()` statement, displaying this value with 3 decimal places.

Solution

```
# Variables:
mays.bats <- 10881
mays.hits <- 3283
mays.doubles <- 523
mays.triples <- 140
mays.home.runs <- 660
mays.bases.on.balls <- 1464
mays.hit.by.pitch <- 44
mays.sacrifice.flies <- 91

# Calculation:
mays.batting.average <-
  mays.hits / mays.bats

cat(
  "Willie Mays batting average:",
  formatC(
```

```

        mays.batting.average,
        format = "f",
        digits = 3
    )
)

```

```
## Willie Mays batting average: 0.302
```

Part (b)

Calculate the number of singles that Mays hit. Report your result using a `cat()` statement.

Solution

```

mays.singles <-
    mays.hits - mays.doubles - mays.triples - mays.home.runs

cat(
    "Willie Mays's singles:",
    formatC(
        mays.singles,
        format = "f",
        big.mark = ",",
        digits = 0
    )
)

```

```
## Willie Mays's singles: 1,960
```

Part (c)

Calculate the total number of outs for Mays. Report your result using a `cat()` statement.

Solution

```

mays.outs <-
    mays.bats - mays.hits

cat("Willie's total outs: ",
    formatC(
        mays.outs,
        format = "f",
        big.mark = ",",
        digits = 0
    ))

```

```
## Willie's total outs: 7,598
```

Part (d)

Calculate the career on-base percentage (OBP) for Mays. Report your result using a `cat()` statement, formatting this value with 3 decimal places.

Solution

```
mays.on.base.percentage <-  
  (mays.hits +  
    mays.bases.on.balls +  
    mays.hit.by.pitch) /  
  (mays.bats +  
    mays.bases.on.balls +  
    mays.hit.by.pitch +  
    mays.sacrifice.flies )  
  
cat(  
  "Willie's on-base percentage:",  
  formatC(  
    mays.on.base.percentage,  
    format = "f",  
    digits = 3  
  )  
)
```

```
## Willie's on-base percentage: 0.384
```

Part (e)

Calculate the number of total bases for Mays. Report your result using a `cat()` statement.

Solution

```
mays.total.bases <-  
  1 * mays.singles +  
  2 * mays.doubles +  
  3 * mays.triples +  
  4 * mays.home.runs  
  
cat(  
  "Willie's total bases:",  
  formatC(  
    mays.total.bases,  
    format = "f",  
    digits = 0  
  )  
)
```

```
## Willie's total bases: 6066
```

Part (f)

Calculate the career slugging percentage (SLG) for Mays. Report your result using a `cat()` statement, formatting this value with 3 decimal places.

Solution**Part (g)**

Construct a pie chart showing the relative proportions of these values for Willie Mays:

- Singles (1B)
- Doubles (2B)
- Triples (3B)
- Home Runs (HR)
- Bases on Balls (BB)
- Hit by a pitch (HBP)
- Outs
- Everything else

Use clockwise rotation for the pie chart, and be sure to provide a main title.

Solution

End of problem 1

Problem 2: Calculating Final Grades

Ashley is registered for graduate credit, and at the end of term she has these results:

- Total problem set points raw score: 58
- Midterm raw score: 69
- Final comprehensive assessment raw score: 76

Part (a)

Calculate the problem set standardized score for Ashley. Report your result using a `cat()` statement, displaying this value with 2 decimal places.

Solution

Part (b)

Calculate the midterm exam standardized score for Ashley. Report your result using a `cat()` statement, displaying this value with 2 decimal places.

Solution

Part (c)

Calculate the comprehensive course assessment standardized score for Ashley. Report your result using a `cat()` statement, displaying this value with 2 decimal places.

Solution

Part (d)

Calculate the preliminary score 1 for Ashley, using a weighted average of the problem set, midterm, and comprehensive course assessment standardized scores. Report your result using a `cat()` statement, displaying this value with 2 decimal places.

Solution

Part (e)

Calculate the preliminary score 2 for Ashley, using a weighted average of the midterm and comprehensive course assessment standardized scores. Report your result using a `cat()` statement, displaying this value with 2 decimal places.

Solution

Part (f)

Based on your results from parts (d) and (e), what is Ashley's final graduate course score? Explain your answer using a few sentences.

Solution

Part (g)

Determine the final course letter grade for Ashley. Explain your answer with a few sentences.

Solution

Part (h)

Miles has the same raw scores as Ashley does, but is registered for undergraduate credit. Report the undergraduate final course score for Miles using a `cat()` statement, displaying this value to 2 decimal places.

Solution

Part (i)

What is Miles' final course letter grade? Explain your answer with a complete sentence.

Solution

End of problem 2

Problem 3: Quadratic Functions

Part (a)

Consider the quadratic function $f(x) = x^2 - 6x + 9$. What is the value of the discriminant for this function? Report your answer using a `cat()` statement, formatting the value to 2 decimal places.

Solution

Part (b)

Based on your answer to part (a), how many solutions exist for the equation $f(x) = 0$? Report your answer with one or two sentences.

Solution

Part (c)

Use the quadratic formula, determine all the roots of the equation $f(x) = 0$. Report each value using a `cat()` statement, formatting the value to 2 decimal places.

Solution

Part (d)

Draw a graph of $f(x)$. The x -values for the plotting region should range from 0 to 6, and the y -values should range from -2 to +8. Include horizontal and vertical reference lines. Draw the curve of $f(x)$, explicitly selecting the line type, width, and color, and then draw in points for any roots of the equation, again explicitly selecting the type, size, and color. Be sure to provide a main title and titles for the x -axis and the y -axis.

Solution

Part (e)

Consider the quadratic function $g(x) = x^2 - 10x + 21$. What is the value of the discriminant for this function? Report your answer using a `cat()` statement, formatting the value to 2 decimal places.

Solution

Part (f)

Based on your answer to part (e), how many solutions exist for the equation $g(x) = 0$? Report your answer with one or two sentences.

Solution

Part (g)

Use the quadratic formula, determine all the roots of the equation $g(x) = 0$. Report each value using a `cat()` statement, formatting the value to 2 decimal places. If there are no solutions, write a short stating that.

Solution

Part (h)

Draw a graph of $g(x)$. The x -values for the plotting region should range from 0 to 10, and the y -values should range from -5 to +5. Include horizontal and vertical reference lines. Draw the curve of $g(x)$, explicitly selecting the line type, width, and color, and then draw in points for any roots of the equation, again explicitly selecting the type, size, and color. Be sure to provide a main title and titles for the x -axis and the y -axis.

Solution

Part (i)

Consider the quadratic function $h(x) = x^2 - 12x + 40$. What is the value of the discriminant for this function? Report your answer using a `cat()` statement, formatting the value to 2 decimal places.

Solution

Part (j)

Based on your answer to part (a), how many solutions exist for the equation $h(x) = 0$? Report your answer with one or two sentences.

Solution

Part (k)

Use the quadratic formula, determine all the roots of the equation $h(x) = 0$. Report each value using a `cat()` statement, formatting the value to 2 decimal places.

Solution

Part (l)

Draw a graph of $h(x)$. The x -values for the plotting region should range from 0 to 12, and the y -values should range from -2 to +20. Include horizontal and vertical reference lines. Draw the curve of $h(x)$, explicitly selecting the line type, width, and color, and then draw in points for any roots of the equation, again explicitly selecting the type, size, and color. Be sure to provide a main title and titles for the x -axis and the y -axis.

Solution

End of problem 3

Problem 4: Grocery Store

A grocery store sells three brands of breakfast cereal:

- Sugar Bomz (SBZ), which costs \$2.49 per box.
- Krispee Yummm!! (KYM), which costs \$3.99 per box.
- Healthy Kale and Tofu (HKT), which costs \$8.99 per box.

The store is currently running a promotional campaign on breakfast cereal by offering a \$1 off coupon to customers.

Note that the coupon is applied to the entire order, not to each box of cereal.

On Saturday, the grocery store had six sales:

Transaction ID	Brand	Number of Boxes	Coupon
1	KYM	2	No
2	SBZ	4	Yes
3	SBZ	6	Yes
4	KYM	1	No
5	HKT	2	Yes
6	SBZ	5	Yes

To calculate the net revenue for each transaction, multiply the price per box by the number of boxes and subtract the coupon amount (if any).

To calculate the total net revenue, add up the net revenues for all transactions.

Calculate the total net revenue for the cereal sales for Saturday.

Report this total net revenue using a `cat()` statement, formatting this value with 2 decimal places.

By the way, we're doing this problem to experience how tedious the calculation is.

It's annoying with just 6 transactions, but think how bad it would be if we had 10,000 transactions!

One of the goals of CSCI S-5a is develop techniques that will allow us to automate this sort of calculation.

Solution

End of problem 4

Problem 5: VWAP

Suppose a stock has five sales in one day:

- First, 1000 shares are sold at a price of \$22.50 per share.
- Next, 200 shares are sold at a price of \$24.00 per share.
- Then 750 shares are sold at a price of \$23.00 per share.
- Then 800 shares are sold at a price of \$24.50 per share.
- Finally, 300 shares are sold at a price of \$24.00 per share.

Let's put this into a table:

Transaction	Number of shares	Price per share
1	1000	\$22.50
2	200	\$24.00
3	750	\$23.00
4	800	\$24.50
5	300	\$24.00

Let's try to summarize these transactions by constructing an “average price per share”.

It doesn't make sense to simply take the average of the price per share for the 5 transactions, because some transactions were for a much larger number of shares.

Presumably, we should weight the price per share more heavily based on the number of shares sold.

This is called a “volume weight average price” (VWAP), because we are constructing a weighted average of the stock prices using the volume (i.e. the number of shares sold) as the weight.

The simplest way to compute the VWAP is to first multiply the number of shares sold by the price per share for each transaction, and then add these together to obtain the total sales amount.

Next, add the number of shares sold for all the transactions to obtain the total number of shares sold.

Finally, the VWAP is the total sales amount divided by the total number of shares sold.

Calculate the VWAP for the 5 transactions, and report your result using a `cat()` statement, rounding to 2 decimal places.

Solution

End of problem 5

Problem 6: Profits

A construction firm has 6 projects in three districts:

Project	District	Costs	Revenue
1	Red	277	362
2	Red	315	397
3	Blue	148	159
4	Blue	206	241
5	Green	483	564
6	Green	388	450

The profit for each project is defined as the revenue for the project minus the cost for the project.

Construct a pie chart showing the relative proportions of the total profits for each district.

Solution

End of Problem 6

Problem 7: The Median Theorem (Extra Credit 1 point)

In Euclidean geometry, any two lines that are not parallel will necessarily intersect in one point.

However, in general three lines will not necessarily intersect in a single point, so when that happens it's "special".

There are many theorems in geometry that deal with three lines all intersecting in a point.

One of the simplest of these theorems is called the "Median Theorem".

A *median* is a line segment from the vertex of a triangle to the midpoint of the opposite side.

The Median Theorem states that the three medians of a triangle always intersect in a single point.

In this problem, we'll use R to create a visualization of this theorem by drawing a triangle and then constructing the three medians.

For best results, I suggest this process:

- First, create an empty plot with no data.
- Next, draw in all the line segments.
- After that, draw the points.
- Finally, add the text annotations.

Part (a): Defining the triangle

Let's start by defining the three vertices of a triangle:

Point	x	y
A	2	3
B	4	18
C	9	12

Create variables to store the x - and y -coordinates of these three points.

Try to use some sort of sensible naming conventions for your variables.

Then create an empty plot with no data, with the x -axis ranging from 0 to 10 and the y -axis ranging from 0 to 20. Be sure to give your graph a main title, and titles for the x - and y -axes.

Finally, draw the triangle ABC . Place points at the vertices, and annotate them with text using the names "A", "B", and "C".

Solution

Part (b): Computing a midpoint

Our next step will be to construct the median from vertex A .

Recall that the median is defined as the line segment from a vertex to the midpoint of the opposite side.

The opposite side from vertex A is the line segment BC , so the median is the line segment from A to the midpoint of BC .

Let's use D to denote the midpoint of the line segment BC , so the median is the line segment AD .

To calculate the x -coordinate of the midpoint of a line segment, take the average of the x -coordinates of the endpoints of the line segment.

That is, take the x -coordinate of the starting point and the x -coordinate of the ending point, add them together, and divide by 2:

$$\text{Midpoint } x\text{-coordinate} = \frac{x_0 + x_1}{2}$$

Similarly, the y -coordinate of the midpoint of a line segment is the average of the y -coordinates of the endpoints of the line segment:

$$\text{Midpoint } y\text{-coordinate} = \frac{y_0 + y_1}{2}$$

Calculate the x - and y -coordinates of the point D i.e. the midpoint of the line segment BC . Save your results in variables, and report each using an individual `cat()` statement, displaying the values with 2 decimal places.

Solution

Part (c): Drawing a median

Using your code from part (a), construct a graph of triangle ABC and add in the median line segment from A to the point D that you calculated in part (b). Include a small point at D and a text annotation.

Solution

Part (d): Drawing another median

In this part, we'll draw another median line.

The side opposite vertex B is AC , so the median line segment from B will end at the midpoint of side AC .

Let E denote the midpoint of side AC , so the median line segment is from A to E .

Calculate the x - and y -coordinates of E i.e. the midpoint AC , and store these values in variables.

Then copy over your code from part (c) for drawing the triangle and add code to draw the median line segment BE .

Solution

Part (e): Drawing the third median

In your graph from part (d), notice that the two medians intersect in a single point.

It would be AMAZING if the third median also went through this point.

In this part, you should copy your code from part (d), and then draw in the third median.

You'll have to figure out what the endpoints are for this third median, and how to calculate them.

When you're all done, the graph should be AMAZING!!

Solution

Part (f): Another triangle

Let's try this again with another triangle:

Point	x	y
A	7	4
B	2	15
C	8	10

Create an empty plot with no data, with the x -axis ranging from 0 to 10 and the y -axis ranging from 0 to 20. Include a main title and titles for the x - and y -axes.

Draw the triangle ABC along with the three medians.

You are welcome and encouraged to copy your code from above and modify it.

If you wrote your code properly, this problem should be very easy, because all you have to do is to alter the values for the coordinates of the three points and then everything else should be defined in terms of these three variables.

Most of your work will be in adjusting the position of the text labels so they look good with these points.

Solution

Part (g): Randomized example generator

You might be wondering if I picked a special triangle where all the three medians just happen to intersect.

Is the Median Theorem really true for ALL triangles?

The only way to definitively answer this question is by a formal mathematical proof (the answer is yes, the theorem is always true).

But another approach to this question is to look at a lot of randomly generated triangles, and if the theorem is true for a sequence of these randomly generated examples then that's a strong indication that it's true in general.

We can use the `runif()` function to generate random values on an interval, which takes three input arguments:

- The number of random values generated, denoted `n`.
- The minimum of the interval, denoted `min`.
- The maximum of the interval, denoted `max`.

To generate one random value on the interval from 0 to 10, we have:

```
runif(  
  n = 1,  
  min = 0,  
  max = 10  
)
```

```
## [1] 8.566778
```

In our example, the x -axis ranged from 0 to 10, so we can generate a random value for the x -coordinate of the point A :

```
A.x <-  
  runif(  
    n = 1,  
    min = 0,  
    max = 10  
  )
```

Similarly, the y -axis ranged from 0 to 20, so we can generate a random value for the y -coordinate of the point A :

```
A.y <-  
  runif(  
    n = 1,  
    min = 0,  
    max = 20  
  )
```

In this problem, use the `runif()` function to generate random values for the x - and y -coordinates of the points A , B , and C .

Then draw the triangle ABC and add the medians.

If you do all of this in a single code chunk, then each time you run the code chunk you'll get a different random triangle, and you should see that the Median Theorem is true in every case.

Hint: if you wrote your code properly in the previous examples, this should be very easy to implement, because once you've generated the random values for A , B , and C then everything else should be defined in terms of these variables so you can just copy your code from above with little modification.

Don't worry about text annotations; it's hard to get the placement right when the points are randomly generated, and anyway all we're doing here is looking to see if the three medians intersect.

Solution