software
education module

Ministry of Science, ICT and Future Planning

KOFAC
Korea Foundation for the Advancement of Science & Creativity
한국과학창의재단

software
education module
10

Sports Statistics

That Bring Victory

Selecting Archers for the Olympics

Statistics for a Winning Shot

Big Data and Soccer

Ministry of Science, ICT and Future Planning

KOFAC
Korea Foundation for the Advancement of Science & Creativity
한국과학창의재단

How to Use This Program

Software is changing the world. The programs installed in computers and apps that make it more convenient to use smartphones are all software. Software is in every part of our lives, so it is difficult to find areas where we are not affected by software.

The state-of-the-art science and technology that we see in the news is also helped by software. In turn, progress in math, science and technology advances software further. As such, math, science and technology are closely related with and cannot be separated from software.

These module series were created through collaboration between experts in related fields and software education, and its suitability for classrooms has been verified. As students follow teachers' direction through each module, they will be able to better understand the world that has been changed by software.

Computational Thinking Map

software
education module

CONTENTS

Intro **Sports Statistics That Bring Victory** -------- 02

PART1 **Selecting Archers for the Olympics** -------- 08

Activity1 **Calculating the Variance of an Archery Score** ------ 12

PART2 **Statistics for a Winning Shot** ---------- 14

Activity2 **Making a Sabermetrics Calculator** -- 20

PART3 **Big Data and Soccer** ---------- 24

Activity3 **Making a Soccer Heat Map** -------- 29

Career Exploration **Big Data Analyst** ---------- 32

Maker Activity **A Marble Rolling Device with Scoring LEDs** -- 34

Sports Statistics That Bring Victory

What are the keys to winning a gold medal at the Olympics? Physical strength, skills, perseverance, and luck! What else? Recently, the importance of statistics in sports has been increasing. Statistics is used in a variety of activities ranging from identifying excellent players to making an optimal strategy to win a game. From here on out, we will look further into the world of "sports statistics," where various kinds of numerical records are kept and analyzed.

Statistics Add to the Pleasure of Watching a Game

Be it soccer, baseball, volleyball, archery, fencing, or track and field, winning or losing depends on players' records that are made during a game. Collecting and analyzing such records might provide a clue to victory. This is the power of "statistics."

Statistics refer to meaningful numbers that allow people to see a social or natural phenomenon at a glance. That is, statistical data are numbers about people, objects, or events, as in the following statement: "As of December 2016, the population of South Korea is 51.7 million, an increase of more than 160 thousand compared to 2015."

Statistics are very useful for guessing what will happen and making a plan for the future because they provide a big picture. This holds true for sports as well. Coaches collect and analyze the statistical records of individual teams and players to establish a winning strategy. Statistics can also contribute to the development of each sport.

Furthermore, statistics add to the pleasure of watching a sports game. Numerical data can help you predict the winner. Admittedly, however, the other team can also beat your prediction. This is why a sports game is called a "drama without a script."

Sometimes, sports statistics serve as a source of inspiration for scientists. Stephen Jay Gould, a representative evolutionary biologist along with Richard Dawkins, is famous for having studied evolution by using baseball statistics.

Gould was interested in the disappearance of the 0.400 batting average in MLB (Major League Baseball) games after Ted Williams' in 1941. He began to examine the season averages of all MLB players starting from 1901 and concluded that the disappearance of the 0.400 average was due to the improvement of all MLB players' skills, rather than the deterioration of the skills of top class batters. The result, suggesting one of the important aspects of biological evolution, helped Gould to elaborate his theory that evolution's drive is towards diversification. A detailed explanation is given in his book *Full House*.

Baseball statistics have been studied in South Korea as well. Scientists conducted the "In-cheon Baek Project" which was named after the only KBO (Korea Baseball Organization) league player to ever achieve a batting average of 0.400. These scientists have greatly contributed to the development of Korean baseball by using a variety of information and statistics provided by the KBO.

Stephen Jay Gould concluded that the disappearance of the 0.400 batting average after 1941 was due to the improvement of the skills of all players. The result supported his theory that evolution's drive is towards diversification.
© Keeton Gale/Shutterstock.com

Sabermetrics to measure

Sports games produce an almost endless amount of statistical data. There are tens of different kinds of official statistics and numerous unofficial ones exist as well.

In the case of baseball, statistics such as batting averages, runs batted in (RBI), home runs, etc. are recorded. In Major League Baseball, a new statistical method called "sabermetrics" has been recently developed in order to evaluate invisible elements such as a player's contributions to his team.

Based on game theory and statistics, sabermetrics aims to understand baseball more objectively. The method originated from the Society for American Baseball Research (SABR) which was founded in the 1970s by Bill James. Sabermetrics was born out of the awareness that traditional statistics such as a batting average and earned run average (ERA) did not truly reflect a player's worth.

For example, a batting average is the number of hits divided by at bats. Walks and hit-by-pitches are not counted in the calculation. Also, the value of a hit is regarded the same as that of a home run, making it hard to measure a player's worth accurately. Thus, a new statistic called "on-base plus slugging" (OPS) was devised. Calculated as the sum of a player's on-base percentage (OBP) and slugging percentage (SLG), OPS helps to evaluate a player's contributions to his team more accurately.

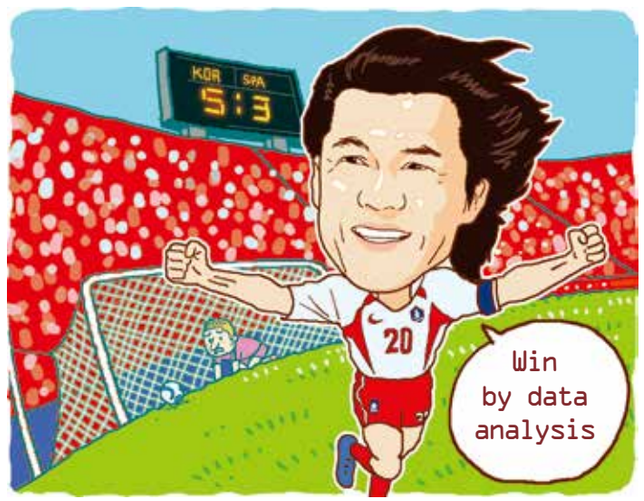
In short, sabermetrics is used to predict a game's result by using accurate statistics that reasonably reflect each player's worth. The Boston Red Sox adopted sabermetrics and in 2004, won the World Series. Statistics are used not only in baseball, but also in many other sports including soccer and archery.

Statistics are also used as an important tool for improving the skills of players. Practice based on various statistical data about oneself and other players can effectively enhance the physical/mental strength and the overall performance of a player.

After sabermetrics founder Bill James was appointed as an Advisor for the Boston Red Sox, the team won its first World Series in 86 years.
© Christopher Penler/Shutterstock.com

A Winning Strategy Supported by Big Data

It was around the time of the 2002 FIFA World Cup when the importance of sports statistics began to be widely recognized in South Korea. Thanks to the effective analysis of the strengths and weaknesses of each team by a performance data analyst with the South Korean national team, South Korea could reach the semi-finals. In the final group game against Portugal, Chong-gug Song contributed to his team's advancement to the round of 16 by effectively blocking Portugal's Luis Figo based on such data. In a penalty shootout in the quarter-final against Spain, the South Korean players utilized data about the Portuguese goalkeeper's positions, reactions, psychological variables, etc. and kicked a ball straight ahead or to the right of the goalkeeper. As a result, South Korea scored all five goals.



In the quarter-final of the 2002 FIFA-World Cup, South Korea scored all five goals in a penalty shootout by utilizing data about the Portuguese goalkeeper.

It used to be thought that all you needed in soccer was being able to run as fast as possible. However, today's soccer has become a "sport of statistics" because data about each player's movement plays an important role in establishing a strategy.

In soccer, various statistics are used such as ball possession percentage, attacking directions, shot directions, attacking positions, corner kicks, and free kicks. Also, a player's ball-possession frequency and duration, speed, heart rate, passing and dribbling directions, etc. are recorded.

Such a vast amount of digital data is called "big data." Their collection and analysis are one of the main strategic aspects of today's sports.

Besides soccer, most other ball games such as baseball and volleyball have dedicated personnel for analyzing data and establishing strategies.

Big data and the German soccer team

IoT refers to the internetworking of the objects around us to enable them to exchange data.

The winner of the 2014 FIFA World Cup, Germany, gained attention with its data-based strategy. In partnership with German big data analytics company SAP, the team's coaches analyzed each player's shots and ball possession time and could reduce the average possession time from 3.4 seconds to 1.1 seconds for speedier passing. Leading up to the semi-final against Brazil, the German team analyzed big data, collected for the last two years, about the rival players' movement patterns, which led to the team's victory.

INTERNET OF THINGS

IoT refers to the internetworking of the objects around us to enable them to exchange data.

Germany's winning of the World Cup has accelerated change in sports strategies, which had traditionally depended on an expert's experience and "gut feeling." The German team's new approach, based on the scientific analysis of big data and the Internet of Things (IoT), proved effective.

The German team used a software called SAP Match Insights. The program receives a vast amount of data via micro-sensors embedded in each player's shin guards, uniforms, and/or soccer balls and analyzes them in real time. In addition, data about each player's movement and ball possession is graphically represented on a screen. This allows for planning a drill that can remedy the shortcomings of individual players and improve the overall efficiency of a team's play. Match Insights is based on two important elements: an IoT network, where numerous sensors communicate with computers, and big data analytics, the process of examining large data sets to find meaningful information.

Big data and the future of sports

Some suggest that in the future, sports might significantly depend on IoT technology. Sensors embedded in a player's uniform, gloves, bat, golf club, racket, helmet, shoes, etc. will generate various data about the player including his/her current location, and the accumulation of such data will allow for evaluating the overall performance of the player. Further development of analytical methods and sensors will help usher in the age of "sports big data," where all the details of each player's motions can be translated into data.

adidas miCoach Smart Soccer Ball has an integrated sensor package that records speed, spin, strike point, etc. when you kick the ball. © adidas

PART 1

Recommended target
Elementary and Secondary Curriculum

Relevant subjects
Middle School Math&Statistics

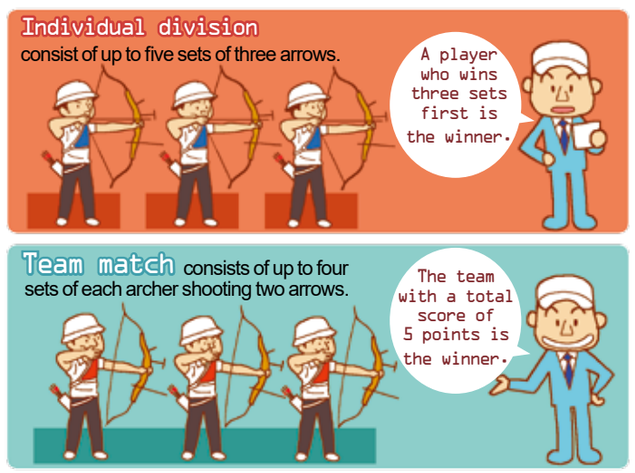
Selecting Archers for the Olympics

Archery is the sport of shooting arrows at a target for accuracy from a set distance. If you were in a position to select individual archers for the Olympics, what kind of players would you choose? Recently, Olympic archery rules have been changed in favor of players who perform consistently well. Then, how do you identify such players?

New archery rules

In the 2016 Rio Olympics, South Korean players won all four gold medals in archery. Prior to that, the World Archery Federation had introduced new archery rules in order to prevent the domination of games by a single country and to make the play more watchable.

Formerly, an individual event in archery was a 12-arrow match consisting of four ends of three arrows, and the player scoring the highest total points was the winner. In the new "set" system, elimination rounds consist of up to five sets of three arrows. A player who wins three sets first is the winner.



New Archery Competition Method

Teams are made up of three archers and a team match consists of up to four sets of each archer shooting two arrows. The team with a total score of 5 points is the winner. If a team wins a set, they receive 2 points; and if the set is drawn, each team receives 1 point.

These somewhat complicated rules are intended to allow for as many unexpected variables as possible. Notwithstanding, South Korea dominated all four events (men's/women's team and individual events) in the Olympics.

Statistics and the selection of archers

In archery, concentration is particularly important. Because archery rules were changed in a way that a single mistake can lead to losing a set, the new rules are in favor of those archers who perform consistently well. Examining each player's points per set can tell who has concentrated better.

The table below shows the points scored by two players, A and B, respectively. Who do you think has better concentration?

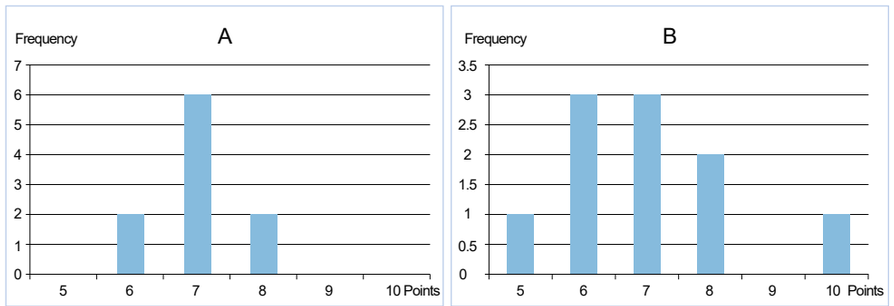
Set	1	2	3	4	5	6	7	8	9	10
Score (A)	8	7	7	6	7	8	7	6	7	7
Score (B)	5	8	10	7	8	7	6	7	6	6

First, let's calculate the mean (average value) of each player's points.

$$\text{Mean (A)} = \frac{8+7+7+6+7+8+7+6+7+7}{10} = 7$$

$$\text{Mean (B)} = \frac{5+8+10+7+8+7+6+7+6+6}{10} = 7$$

The two mean values are the same. Because it is difficult to compare the distributions of the points by using mean values alone, the following histograms have been made.



Archery score histograms

Player A's points are closer to the mean, while player B's points are spread out over a wide range of values.

Measures of dispersion in the distribution of a player's points are useful for evaluating the consistency in the player's performance.

In the above histograms, player A's points are closer to a mean value of 7, while player B's points are spread out over a wide range of values despite the two players sharing the same mean value.

In statistics, the extent to which a distribution is spread out is called dispersion. Player A's points have low dispersion, whereas player B's points have high dispersion.

Calculating dispersion

There are various measures of dispersion. You can simply show a range from the lowest to the highest value, or you can calculate how far each data element is from a mean value as discussed above. "Variance" is one of the most representative measures of dispersion.

The table below shows the deviations of each player's points. A deviation refers to the difference between a number in a set and the mean of the set. Each deviation is calculated by subtracting the mean from each number. Because a deviation can be either positive or negative, the absolute value of each deviation is used for comparison. The larger the absolute value of the deviation of a data element is, the farther the element is from the mean.

Set	1	2	3	4	5	6	7	8	9	10	Sum	Mean
Score (A)	8	7	7	6	7	8	7	6	7	7	70	7
Deviation/ AV	1/1	0/0	0/0	-1/1	0/0	1/1	0/0	-1/1	0/0	0/0		
Deviation^2	1	0	0	1	0	1	0	1	0	0		
Score (B)	5	8	10	7	8	7	6	7	6	6	70	7
Deviation/ AV	-2/2	1/1	3/3	0/0	1/1	0/0	-1/1	0/0	-1/1	-1/1		
Deviation^2	4	1	9	0	1	0	1	0	1	1		

The distribution of the points scored by player A and player B

While a deviation can be used to show how far a data element is from the mean, it cannot indicate how all the elements are spread out. In order to know that, variance is used. A variance is calculated by dividing the sum of squared deviations by the number of data elements.

Using the squared deviations in the above table, the variances of the points of player A and player B are calculated in the following way.

$$\text{Variance (A)} = \frac{1^2+0^2+0^2+(-1)^2+0^2+1^2+0^2+(-1)^2+0^2+0^2}{10} = \frac{2}{5}$$

$$\text{Variance (B)} = \frac{(-2)^2+1^2+3^2+0^2+1^2+0^2+(-1)^2+0^2+(-1)^2+(-1)^2}{10} = \frac{9}{5}$$

As is shown above, player A's variance is lower than player B's. A lower variance means that data elements are closer to its mean.

Let's return to the topic of selecting a player. Of the two players, player A is likely to be chosen because the player's points have lower variance, indicating a more consistent performance than player B's.

In short, you must calculate not only an average but also measures of dispersion such as variance in order to understand the characteristics of a data set. Measures of dispersion are widely used, for example, for student score distributions and the annual temperature distributions of a region.

Although the average annual temperatures of Minneapolis in the U.S. and Plymouth in England are almost the same (about 13°C), the difference between the highest and the lowest temperature in Minneapolis is much larger than that of Plymouth. Therefore, it is necessary to know the temperature distribution of a region in order to understand the climate characteristics of the region.

Although the average annual temperatures of Minneapolis (above) and Plymouth (below) are almost the same, the two regions' annual temperature distributions are different.

Activity 1

Calculating the Variance of an Archery Score

In this activity, you will play an archery game with your friends and record the points scored to calculate their dispersion. Also, you will select a player who performed consistently well.

What to prepare Smartphone, pen

Activity

1 Install Archer World Cup III in your smartphone.

2 Prepare a pen.

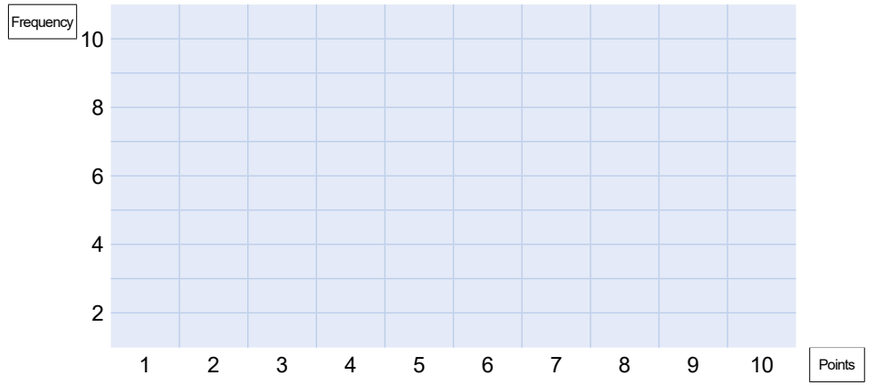
3 The following rules are applied:

- ► The game is made up of five sets, and three players take turns in shooting each of three arrows per set.
- ► The player with the highest total points in a given set receives a set score of 2 points; the next highest, 1 point; and the lowest, 0 points
- ► The player with the highest total set score is the winner.

4 Play the game and record each point and set score in the table below.

Name		Set 1	Set 2	3Set 3	Set 4	Set 5	Total set score
	Points						
	Set score						
	Points						
	Set score						
	Points						
	Set score						

5 Make a histogram of the distribution of your points. The x-axis indicates the points you scored and the y-axis indicates the frequency of each point.



6 Using the score table in 4, calculate each mean and variance.

Name		Set 1	Set 2	Set 3	Set 4	Set 5	Mean
	Points						
	Deviation						
	Deviation^2						
	Variance						
	Points						
	Deviation						
	Deviation^2						
	Variance						
	Points						
	Deviation						
	Deviation^2						
	Variance						

※ You can use a calculator app on your smartphone.

7 Compare the performances of the players in terms of mean, total set score, and variance.

Player with the highest mean	Player with the highest total set score	Player with the lowest variance

PART 2

Recommended target
Elementary and Secondary Curriculum

Relevant subjects
Equation, Grade 1(Middle school Math)

Statistics for a Winning Shot

Professional baseball requires an enormous amount of expenditure, including player salaries, ballpark maintenance costs, and publicity expenses. Thus, each team tries its best to win, which requires excellent players. Team owners carefully consider which players to hire and whether their price tags are reasonable. In the field, managers have to decide which players to put in the lineup. Such decision-making used to be based on the intuition of an owner or a manager. However, today's decision-making is usually based on statistics.

What is sabermetrics?

Baseball is called a "sport of statistics." Most of the in-game records are expressed in numbers such as a batting average, earned run, and on-base percentage. Sabermetrics* refers to a method of evaluating individual teams and/or players based on objective statistical data. In other words, a player's specific in-game records are used for evaluation, instead of subjective impressions about how he plays.

It was Billy Beane, then general manager of the Oakland Athletics, who first began to use sabermetrics to manage a team. He scouted new players based on the statistical analysis of their season records. As a result, the Athletics set a record of winning 20 consecutive games. Beane's historic success story was made into the 2011 film *Moneyball*.

*** Sabermetrics**
The term is derived from the acronym SABR (Society for American Baseball Research) founded by Robert Davis and the word metrics (meaning "measuring").

Sabermetrics refers to a method of evaluating individual baseball teams and/or players based on objective statistical data.
© Richard Paul Kane/Shutterstock.com

Sabermetrics is not used in South Korean baseball because player data has not been recorded enough to apply the method. Besides, analytical methods that reflect the real-world situations of Korean baseball have yet to be developed.

Representative baseball statistics

In baseball, various in-game statistics are recorded. The following are some important baseball statistics:

① Batting average (AVG) | the percentage of at bats (AB) in which a batter gets a hit. A batter with a high AVG is one who has many successful hits. Walks and hit-by-pitches are not counted in calculating an AVG. Nor are sacrifice hits and flies that bring a runner home.

[AVG = (1B+2B+3B+HR]÷AB*]

② Slugging percentage (SLG) | the rate of power hits for a batter. An SLG is calculated by taking a batter's total bases divided by at bats.

[SLG = (1B×1+2B×2+3B×3+HR×4)÷AB]

③ On-base percentage (OBP) | the probability of a batter reaching base for any reason other than a fielding error, fielder's choice, or uncaught third strike. Batters with a strong OBP, such as Shin-soo Choo of the Texas Rangers, are highly recognized in modern baseball.

[OBP = (1B+2B+3B+HR+BB+HBP)÷PA*]

④ On base plus slugging (OPS) | the sum of a batter's OBP and SLG. OPS has been recently regarded as a baseball statistic that reflects a batter's performance more accurately than his batting average.

[OPS = OBP+SLG]

In baseball, various in-game statistics are recorded and used to evaluate individual players and/or teams.
© Photo Works/Shutterstock.com

*** PA vs. AB**
Plate appearances (PA) are the number of times a player completed a turn batting, while at bats (AB) refer to PA excluding sacrifice hits and flies, walks, and hit-by-pitches.

Expected runs scored and allowed

Like any other sport, chance or luck play a big part in baseball. A runner on first base with two outs can be driven home by a hit, whereas bases loaded with no outs* can produce zero runs. A game can be won with 10 hits allowed but allowing no runs, whereas a game can be lost with just one hit allowed.

Then, what is the use of sabermetrics? "Statistics" and "luck" may look contradictory to each other. However, the statistical comparison of expected values and real values allows for the analysis of chance or luck. A representative statistic in this context is "Base Runs."

Developed by David Smith in 1990, Base Runs is a statistically expected value. More specifically, it measures the expected runs scored/allowed based on the batting and pitching records of a team.

Expected runs scored and expected runs allowed are calculated by the following formulas, respectively.

$$\text{① Expected runs scored} = A \times \frac{B}{(B+C)} + D$$

A = H+BB+HBP-HR-(IBB*×0.5)

B = {1.4×TB*–0.6×H–3×HR+0.1×(BB+HBP–IBB)
+0.9×(SB-CS–GIDP)}×1.1

C = AB–H+CS+GIDP

D = HR

$$\text{② Expected runs allowed} = A \times \frac{B}{(B+C)} + D$$

A = H+BB+HBP–HR-(IBB×0.5)

B = {1.4×TBA*–0.6×H–3×HR+
0.1×(BB+HBP–IBB)+0.9×(SB–CS)}×1.1

C = IP×3

D = HR

*** Bases loaded with no outs**

A situation in which runners are on first, second, and third base. This presents a great opportunity for scoring at least one run.

*** Intentional walk (IBB)**

A walk issued to a batter when a pitcher deliberately throws far outside the strike zone.

*** Total bases (TB)**

The number of bases gained by a batter through his hits.

*** Total bases allowed (TBA)**

TB allowed by a pitcher

Pythagorean expectation

Invented by Bill James, the founder of sabermetrics, Pythagorean expectation estimates an expected win ratio based on runs scored and allowed by a team. The name is derived from the formula's resemblance to the Pythagorean theorem*. The following is the basic formula.

$$\text{Win ratio} = \frac{\text{Runs scored}^2}{(\text{Runs scored}^2 + \text{Runs allowed}^2)}$$

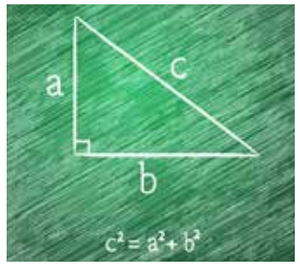
The above formula is based on MLB games. In South Korea, the following adapted formula was made based on the statistical analysis of the KBO data from 1991 to 2012.

$$\text{Adapted win ratio} = \frac{\text{Runs scored}^{1.87}}{(\text{Runs scored}^{1.87} + \text{Runs allowed}^{1.87})}$$

Depending on a team's characteristics, the actual winning percentage can be different from the Pythagorean winning percentage. The above formulas suggest that the more runs allowed, the lower the win ratio becomes. However, if a team keeps winning despite allowing many runs, the team's actual win ratio will be higher than the Pythagorean ratio. Besides, a team with excellent pitchers is likely to have a higher actual win ratio because they can win by allowing fewer runs than they scored.

*** Pythagorean theorem**

A statement that the square of the hypotenuse is equal to the sum of the squares of the other two sides, as is shown in the figure below.



WAR, an indicator of a player's contributions to his team

Wins Above Replacement (WAR) is one of the most frequently used statistics in modern baseball. WAR shows how many more wins a player is worth than a replacement-level player. In 2010, for example, the WAR value of Dae-ho Lee of KBO's Lotte Giants was 8.76, suggesting that his team won roughly nine more games than would be expected if he were substituted by a replacement-level player.

WAR is calculated by using the comprehensive records of a player and is adjusted for league and ballpark variables. Therefore, it is pos-

+ A player's worth by WAR

0-1 WAR Scrub
1-2 WAR Role player
2-3 WAR Solid starter
3-4 WAR Good player
4-5 WAR All-Star
5-6 WAR Superstar
6+ WAR MVP

sible to compare Babe Ruth's WAR value of 14.0 (1921) and Shin-soo Choo's WAR value of 7.3 (2010).

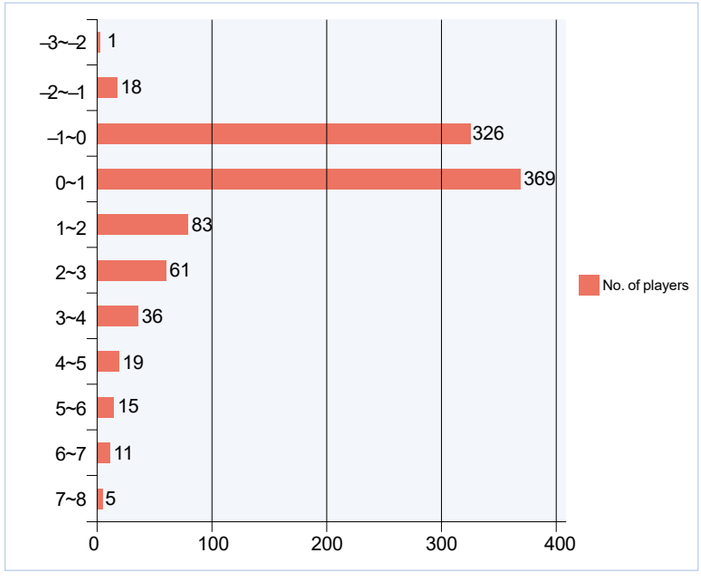
Because the calculation of WAR requires various kinds of data and their adjustments according to situations, it cannot be represented as a single formula. Thus, WAR is considered to be one of the most difficult baseball statistics to calculate.

Setting criteria is one of the most important parts of WAR calculation. In the context of WAR, a "replacement-level player" is defined as a player who

- Has a below-league-average performance;
- Is easily available in the trade market; and
- Badly influences his team's performance when he is included in the lineup.

In the above histogram, the WAR values of most of the MLB players are between –1 and 1. A WAR value of 4 and above means an excellent player within the top 6%.

As of 2010, the WAR values of most of the MLB players were between –1 and 1. A WAR value of 4 and above means an excellent player within the top 6%.



The distribution of the WAR values of the MLB players as of 2010

WAR can be said to be a statistical answer to the following important question: "Who is the most valuable player?" In the past, players with strong records in batting statistics such as home runs, RBIs*, and hits were highly recognized. However, WAR measures a player's contributions to his team by considering all elements that affect the games, including his team, league, year of the league, and his position. Therefore, a player with an exceptional fielding performance can be highly recognized as well. For example, you can compare the contributions of a power-hitting first baseman and those of a near-perfect shortstop.

Now, let's evaluate the two players in the table below. Choose the player who has contributed to his team more and write down the reason for your choice.

*** Runs batted in (RBI)**

Runs scored due to a batter's action

	HR	RBI	AVG	OPS
Player A	15	30	0.346	1.089
Player B	10	21	0.233	0.922
Better player				
Reason				

Activity ❷

Making a Sabermetrics Calculator

With the introduction of sabermetrics, players who used to be underestimated by standard statistics such as AVG, HR, and RBI were able to be revaluated. In particular, OPS (the sum of SLG and OBP) is a representative sabermetrics statistic that effectively indicates a player's worth. In this activity, let's make a program for calculating OPS easily.

What to prepare A computer in which Scratch (version 2.0) is installed

Activity

1 Mission

Create a project that meets the following conditions.

Conditions
① The program calculates AVG, SLG, OBP, OPS, AB, and PA simultaneously.
② Variables are used to run the program.

2 Project design

① Write down formulas for calculating SLG, AVG, OPS, and OBP.

Statistics	Formulas
SLG	
AVG	
OPS	
OBP	

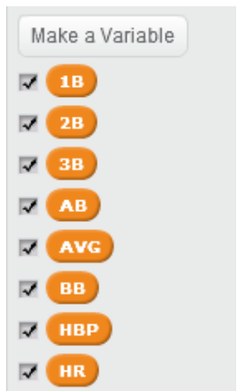
② Draw an algorithm for the baseball calculator project.

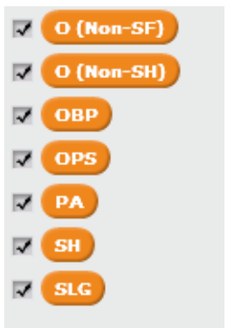
3 Programming

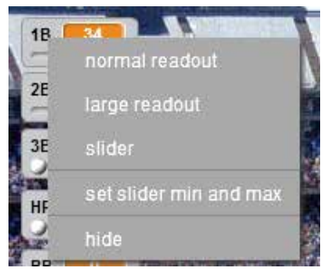
① Run Scratch and select a ballpark image as a backdrop.

② Add a sprite in the "New sprite" menu and place it in an appropriate position.

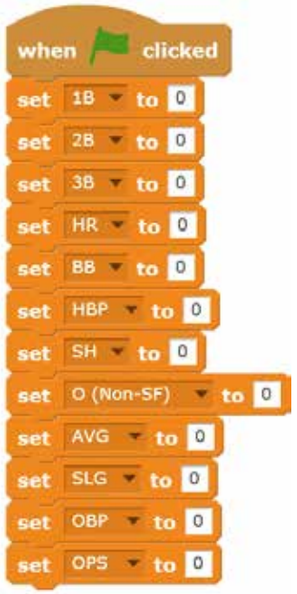
③ Many variables are needed to make a baseball calculator. In the "Data" menu, click on the "Make a variable" button. Give each variable a name: 1B, 2B, 3B, OPS, HBP, BB, O (Non-SH), SLG, OBP, PA, AB, AVG, HR, SH, and O (Non-SF), where O stands for outs; SH, sacrifice hits; and SF, sacrifice flies. Right click on the orange part of the variables and select "Slider".



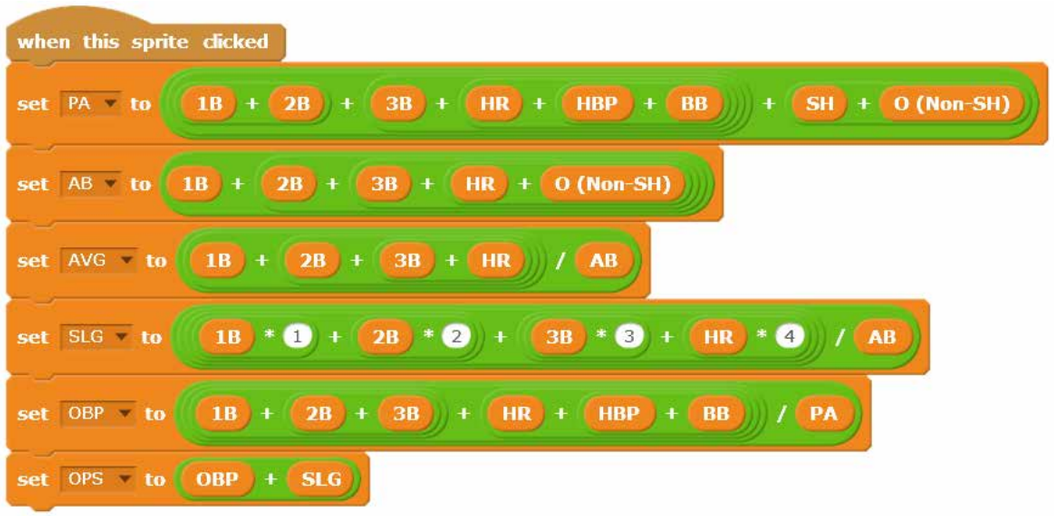




④ Assemble the blocks so that all variables will be reset when you click on the green flag.

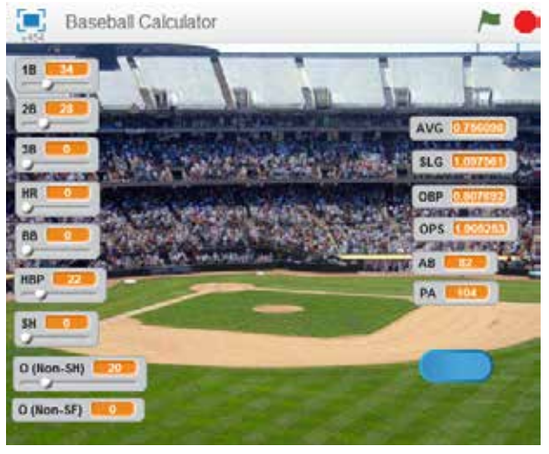


⑤ Assemble the blocks so that PA, AB, AVG, SLG, OBP, and OPS will be calculated when you click on the “Calculate” button.



⑥ Run the program to check for an error. If there is any, correct the error.

⑦ End result (screenshot)



4 Choose either of the following projects to change the program

① Make a calculator that has a Q&A format.

② Make a calculator for a different sport.

5 Summing up

Summarize what you felt through this activity.

PART 3

Recommended target
Elementary creative experience activities

Relevant subjects
Practical course

Big Data and Soccer

Soccer is one of the most popular sports in the world. A vast number of people watch the FIFA World Cup and the UEFA Champions League games, to name a few. Compared with other sports, chances are higher in soccer that an underdog can beat a strong team, which makes watching a soccer game very exciting. By the way, is it possible to statistically analyze soccer games, as is the case with sabermetrics in baseball? In the past, that might have sounded like a dream. In today's age of big data, however, the "dream" has come true.

A hard sport to quantify

What kinds of soccer data are there? Compared to other sports, soccer has many variables and is a sport whose results are hard to predict. When sports bettors place their wagers on teams, the odds per team are decided. Usually, the odds for strong teams are lower than those for underdogs because strong teams are more likely to win.

In soccer, the probability of a strong team beating an underdog is only 53%. This is because soccer has many variables compared to other sports.
©AGIF/Shutterstock.com

In soccer, the probability of a strong team beating an underdog is only 53%, which suggests that both a strong team and an underdog have about a fifty-fifty chance of winning due to numerous variables.

Just like baseball, soccer does have statistics to evaluate individual teams and players, such as goals, assists, yellow/red cards, and corner kicks. Unlike baseball, however, a relatively small number of statistics have been used in soccer. Because each half of a soccer game runs continuously for 45 minutes, it is very difficult to manually keep a record of all the game details. Thus, soccer was regarded for some time as a sport that does not produce many kinds of data by its nature.

Until recently, a relatively small number of soccer statistics have been used, such as goals, assists, yellow/red cards, corner kicks, etc.
©Paolo Bona/Shutterstock.com

Recently, however, things have changed with regard to soccer statistics. In addition to traditional statistics, new categories, such as pass success percentage, total run distance, dribble distance, and key pass (a pass given by a player to a teammate, who is then able to take a shot at goal), began to be used thanks to the development of data collection and analysis techniques.

Whereas baseball is called a sport of statistics, soccer is regarded as one of the most difficult sports to keep a record of the play. Assuming that you are an official scorer, write down the differences between soccer and baseball.

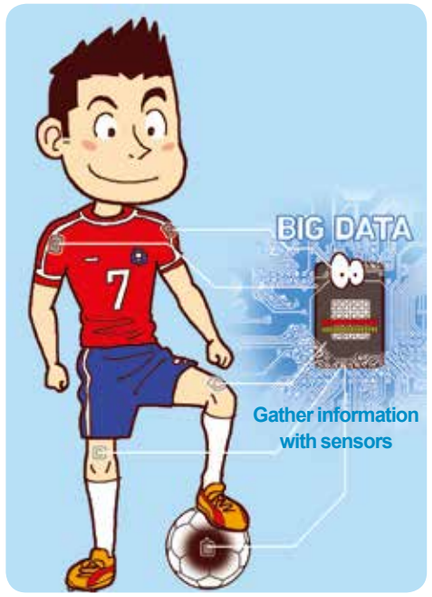
Soccer	Baseball

What kind of technique do you think will help to keep a record of every detail of a soccer game? Assuming that you are an official scorer, describe a necessary technique. It is OK if the technique has not been developed yet.

Various in-game data about German national soccer player Philipp Lahm.
Big data technology allows for the effective analysis of soccer data. © SAP

Sensors that collect big data

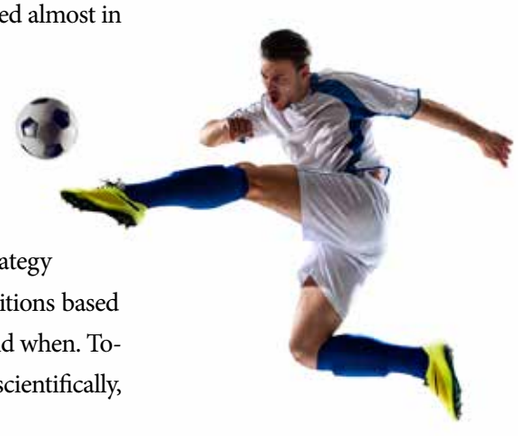
It was the German national soccer team that made soccer another sport of statistics. Leading up to the 2014 FIFA World Cup in Brazil, the coaches of the German team trained their players in a special way. During practice, each player wore small devices with embedded sensors so that the German IT company SAP could monitor their movement and health status.



SAP specializes in big data analytics. Big data refers to large, complex data sets that are created in a digital environment. They include not only numbers but also text and video data. Because data streams are usually collected by computers at short time intervals, a vast amount of data is collected, which is why big data experts are needed to analyze them and extract meaningful information.

Via the sensors attached to the players, SAP was able to measure their training load, breathing rate, heart rate, etc., and collected as many as about 12,000 raw data per minute. As a result, a single training session produced hundreds of thousands of raw data about each player. The coaches leveraged SAP's analytics about their players in strategy setting and individual practice planning, which led to winning the World Cup.

Storing and analyzing soccer data is also supported by camera tracking technology. In Europe, a huge amount of raw data, including those about the movement, distance and speed of each player, the distance between players, and the direction of the ball, is produced by 16 cameras placed around each soccer field. Based on such raw data, more than 2,000 pieces of analyzed information is provided almost in real time.



Pep Guardiola, manager of Manchester City, actively applies such real time data to games. During a half-time break, he checks the data created during the first half of the game with his players and discusses whether they should change their current strategy or not. In the past, managers determined a player's conditions based on what they saw and felt, and decided who to replace, and when. Today's managers, however, are able to establish strategies scientifically, based on accurate statistics and visual data.

Big data technology allows for the effective analysis of each player's movement, and the resultant data pieces are provided in real time for setting a strategy.
© Jefferson Bernardes/Shutterstock.com

Heat maps for identifying a player's movement pattern

A heat map is an analysis tool that provides a graphical representation of data. In soccer, heat maps are used as an indicator of the scope and frequency of a player's movement, helping to identify his conditions and in-game situations. In the past, heat maps were drawn manually by marking, at a certain time interval, a player's position and integrating the marks.

These days, however, computers draw heat maps thanks to the development of technology. On a heat map, a player's movement is represented by different colors and depths. The color gets redder in areas where a player's presence increases. The time interval of a heat map's data collecting can be set as desired. A shorter interval means more data collected, which increases accuracy but also increases calculating time. On the contrary, a longer interval means less accuracy but less calculating time is needed.

Based on a heat map produced in real time, a manager can see whether a player is moving according to his strategy or can decide who to replace by considering each player's amount of activity.

A heat map is an indicator of a player's movement. The color gets redder in areas where the player's presence increases.
© www.whoscored.com

Activity ❸

Making a Soccer Heat Map

Until somewhat recently, collecting and analyzing soccer data was a very difficult task. Heat maps are one of the analytical tools used in soccer. A traditional heat map was drawn by marking, at a certain time interval, a player's position and integrating the marks. The color of a heat map gets deeper in areas where a player's presence increases and lighter in areas where this is not the case. Heat maps are useful for easily identifying each player's movement. In this activity, let's make a simple heat map program, in which a mouse cursor takes the role of a soccer player.

What to prepare A computer in which Scratch (version 2.0) is installed

Activity

1 Mission

Create a project that meets the following conditions:

Conditions
① A soccer field image is used as a backdrop.
③ The color changes according to the movement of the mouse.
③ A "stamp" effect is used.

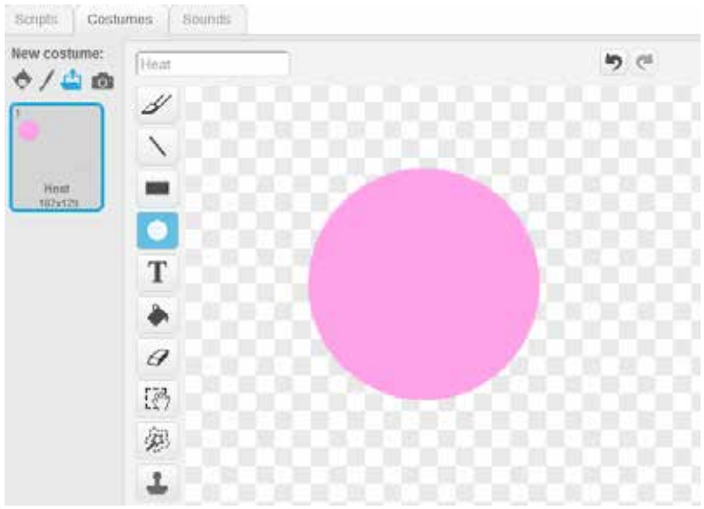
2 Project design

Draw an algorithm about how to execute the program, including movement directions.

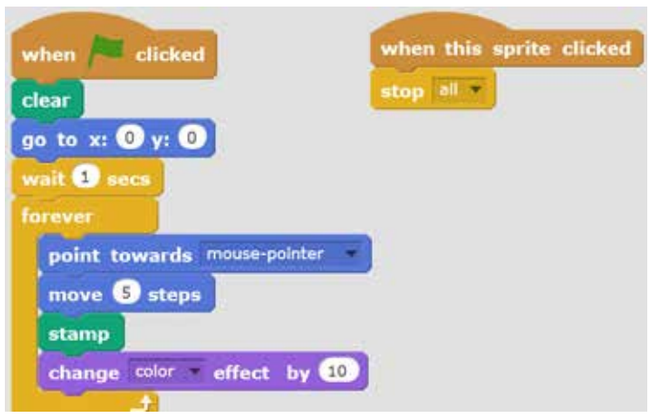
3 Programming

① Create a soccer field backdrop. Because there are no soccer field image files in Scratch, you have to upload your own file.

② In order to add a new sprite, click on the “Paint your own sprite” button in the “New sprite” panel and paint your “Heat” sprite.



③ Assemble the blocks so that the existing heat map will be reset when you click on the green flag. Make the script stop when you click on the “Heat” sprite again.



④ Run the program to check for an error. If there is any, correct the error.

⑤ End result (screenshot)

4 Choose one of the following projects to change the program

① Make a heat map program that is executed by using the arrow keys.

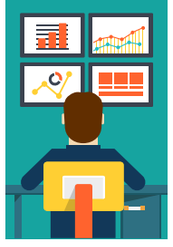
② Design your own unique heat map and apply it to your program.

③ Make a heat map program that has sound effects.

5 Summing up

Summarize what you felt through this activity.

Career Exploration



Big Data Analyst

The development of IT and the explosive increase of data have ushered in the "age of big data." Meaningful information extracted from a vast amount of data serves as an important basis for scientific and rational decision-making. Big data experts collect and manage a large number of data sets that seem to have nothing in common. They dig into big data to search for various kinds of information such as the behavioral patterns of people or expected economic situations. In order to do this, you must be knowledgeable about statistics, business consulting, data analysis techniques, etc., Along with the further development of computers, mobile technology, the Internet of Things (IoT), SNS, etc., the big data market will continue to grow, and so will a demand for big data experts such as big data analysts.

What do they do?

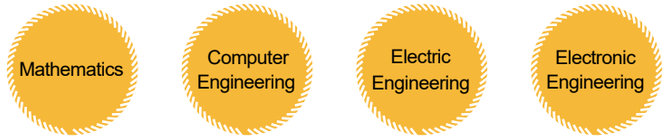
- They design methods for collecting data created in various fields
- They construct data storage architecture; design data flow; and manage databases
- They analyze and process big data to provide meaningful information that brings convenience to daily life and/or research activities

Occupations related to big data analytics

Data science	Data mining expert, information manager, sports data analyst
IT	Database manager, statistics analyst, software developer
Government and public agencies	Government statistician, researcher in statistics, mathematics, natural science, or satellite development

Related majors

Various skills are needed in order to process a vast amount of data from different fields such as social/natural science and economics. Artificial intelligence (AI), which



has recently surprised the world with its possibilities, is based on big data technology. Because fast computers are essential for processing massive data streams, you need to be skilled in computer hardware. Mathematical knowledge and insights are also needed because statistical methods are used in data processing.

Required aptitude

- Competence in programming languages is a basic requirement for dealing with databases. Theoretical and practical experiences with databases are also required.
- Big data is collected in forms that make it difficult to identify commonalities or regularities. Therefore, you are required to have mathematical insights and analytical skills as well as the capability to solve problems with algorithms.
- A learning attitude is required, as big data is an area where new technologies are being rapidly developed.
- Interpersonal and communication skills are also necessary because you will often collaborate with other technicians or experts.

Expert interview

Statistics aims to visualize a large data set or search for patterns in data.
For example, it would be more efficient to present a representative value such as mean than simply to list the math scores of all students of a school. Statistical methods are also useful for identifying the patterns of climate change in daily weather data or the efficacy of a new drug. In other words, statistics is like assembling Lego bricks to construct an object.
Recently, the number of data "bricks" have increased and their shapes have become more complicated, making it impossible or inefficient to apply the existing assembling methods. In order to solve these problems, I would like to expand the scope of existing statistical methodologies. For now, I am learning basic statistical theories and programming, after which I will begin to conduct my own research.

Seong-oh Park / Ph.D. candidate, Dept. of Statistics, Seoul National University

Maker Activity

A Marble Rolling Device with Scoring LEDs

Written by
Dong-man Kim (Incheon Buhyun Elementary School)

Product overview

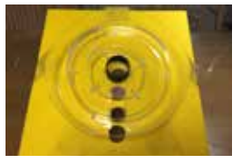
Product function

A device for a marble rolling game where a player's points are displayed by LEDs

How it works

Points are assigned to each area that has a hole. A switch is installed underneath a tube connected to each hole. When a marble drops into a hole, the switch is pressed and the corresponding LED is turned on.

Product picture



Front

Back

Clip switch

End result

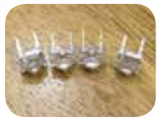
Product structure

Clip switch
A device that sends an input signal to the Arduino Uno board

Entry
A programming language for sending commands to the Arduino Uno board

\+

Device framework consisting of a foam board, plastic bottle cut-outs, and an OHP film



LED
Output devices that receive commands from the Arduino Uno board

Production overview

- **Production time** : 3.5 hours
- **Needed materials and tools**

▲ Arduino Uno board

▲ LEDs (4 colors)

▲ Wires

▲ OHP film

▲ Plastic bottles (4 sizes)

▲ Foam board

▲ Cutter

▲ Transparent tape

▲ Cable ties

▲ Clips

▲ Awl

▲ Wire mold

Key production principles

Connecting an Arduino Uno board to display scores with LEDs when the switch is pressed.

When a marble drops into a hole, the switch is pressed and the Arduino Uno board receives an input signal to turn on an LED. The four LEDs have different colors which correspond to each point scored. Power is supplied to the LEDs via the Arduino board.

What requires attention in production

❶ Make 3D scoring areas so that they are distinct from each other.
❷ Use marbles that are heavy enough to press the switch.
❸ Bend the lower part of the marble track more than 30 degrees upward so that it can serve as a schanze.
❹ Make the switches in a way that the gap between the clips is not too wide.

Required knowledge and functions

❶ How to use an Arduino Uno board and how to make a program by using Entry
❷ How to make a switch and understanding of how LEDs work
❸ The safe use of tools such as a cutter, awl, and cable ties

SW Education Module Textbook

- ❶ Artificial Intelligence
- ❷ Driverless Vehicles
- ❸ IoT(Internet of Things)
- ❹ Virtual Reality
- ❺ CRISPR
- ❻ Space Launch Vehicles
- ❼ Natural Disasters
- ❽ Smart Medicine
- ❾ Game Engines
- ❿ **Sports Statistics**

Problem-Solving Activities for Computational Thinkers ❿ Sports Statistics

Sports Statistics That Bring Victory

Issued on February 23, 2017
Issued by Ministry of Science, ICT and Future Planning, Korea Foundation for the Advancement of Science and Creativity
Written by Chang Ho Ahn
Planned by Ji Young Park, Jong Keol Son, Hye Ja Kim (MSIP)
Yun Jeong Kim, Se Man Park, Oksu Hong, Hannah Jung (KOFAC)
Reviewed by Hye Jin Nam, Beyong Rae Han, Yeong Jae Jang

Published by Doo Hui Kim
Publisher Dongascience Inc.
Responsible editor Kyeong Ae Chang

Inquiries Office of Software Education, Korea Foundation for the Advancement of Science and Creativity
Email swedu@kofac.re.kr
Phone (02)559-3841~9
Address (135-867) 602 Seolleung-ro, Gangnam-gu, Seoul (Samreung Building, Samseong-dong)
Homepage www.kofac.re.kr