Chapter 3

Will Women Run Faster than Men in the Olympics?

The question posed in the title of this chapter cannot be answered for certain. However, students can develop a plausible answer to it by reasoning with data and applying important connections to other aspects of their growing awareness of mathematics.

The Context

This investigation presents students with available Olympic data and asks them to develop a statistical model that can help answer the question. The data were collected at each of the modern Olympics that included the 200-meter dash. Students think about how this observed data could be analyzed. They are often amazed by the patterns that emerge and the implications. The dialogue in this chapter is intended to demonstrate some typical student reasoning and sense making about comparing data sets collected over time.

The Big Statistical Ideas

The discussions in this chapter focus on interpreting and developing various equations of least-squares regression lines. It is important that students think and reason about regression lines as summary representations of data that can be used to make conjectures about the data. Although understanding all the aspects of the specific techniques is not critical for deriving the equation of the regression line, students should examine the line's fit to a scatter plot of the data, describe how the data are distributed in the plane about a line of fit, and draw appropriate conclusions to address questions posed about the data, such as the Olympic comparison question posed for the investigation. In the end, students may conclude that the question cannot be answered with the data at hand, but the process of reaching that conclusion is important in developing their reasoning and sense making skills. Table 3.1 identifies the key element and summarizes the general reasoning habits and the specific statistical reasoning habits of mind that are involved in this investigation.

Table 3.1
Key Element and Habits of Mind in the Olympic Exploration

Key Element: Analyzing Data

Habits of Mind

Analyzing a problem

Looking for patterns and relationships by—

- describing overall patterns in data;
- looking for hidden structure in the data;
- making preliminary deductions and conjectures.

Monitoring one's progress

Evaluating a chosen strategy by—

- evaluating the consistency of an observation with a model;
- applying the iterative statistical process to the investigation.

Seeking and using connections

Connecting different representations by—

- identifying common components of analyses (e.g., standardization);
- understanding the sensitivity of an analysis to various components;
- connecting conclusions and interpretations to the context.

Reflecting on one's solutions

Checking the reasonableness of an answer by—

- determining whether a conclusion based on the data is plausible;
- justifying or validating the solution or conclusion.

Introduction to the Task

The Summer Olympic games are generally held every four years. They provide an opportunity for athletes from all over the world to compete. As expected, this level of competition results in the best athletes demonstrating their abilities to run, jump, throw, wrestle, box, and excel in many other individual and team sports. The track and field events are particularly noteworthy, as the Olympics provide a prime-time audience with access to competitions not normally showcased. The running events are frequently won or lost by a fraction of a second. The gold medal times for the 200-meter dash are listed in table 3.2 and represent all of the modern-day Olympic winning times from 1900 to 2004 for this race.

This investigation calls on students to examine the data in the table and consider the following question: Can you use these data to predict the future times of men and women in this event? Do you think women will ever run faster than men, and if yes, when? Construct an argument to support your answer to this question by using data.

Table 3.2
Times for the Olympic 200-Meter Dash

Year	Male	Time (in seconds)	Female	Time (in seconds)
1900	Walter Tewksbury, USA	22.2		
1904	Archie Hahn, USA	21.6		
1908	Robert Kerr, Canada	22.6		
1912	Ralph Craig, USA	21.7		
1920	Allan Woodring, USA	22.0		
1924	Jackson Scholz, USA	21.6		
1928	Percy Williams, Canada	21.8		
1932	Eddie Tolan, USA	21.12		
1936	Jesse Owens, USA	20.70		
1948	Mel Patton, USA	21.10	Fanny Blankers-Koen, NED	24.40
1952	Andy Stanfield, USA	20.81	Marjorie Jackson, AUS	23.89
1956	Bobby Morrow, USA	20.75	Betty Cuthbert, AUS	23.55
1960	Livio Berruti, ITA	20.62	Wilma Rudolph, USA	24.13
1964	Henry Carr, USA	20.36	Edith McGuire, USA	23.05
1968	Tommie Smith, USA	19.83	Irena Szewinska, Poland	22.58
1972	Valeriy Borzov, USSR	20.00	Renate Stecher, GDR	22.40
1976	Don Quarrie, JAM	20.23	Barbel Eckert, GDR	22.37
1980	Pietro Mennea, ITA	20.19	Barbel Wockel (Eckert), GDR	22.03
1984	Carl Lewis, USA	19.80	Valerie Brisco-Hooks, USA	21.81
1988	Joe DeLoach, USA	19.75	Florence Griffith-Joyner, USA	21.34
1992	Mike Marsh, USA	20.01	Gwen Torrence, USA	21.81
1996	Michael Johnson, USA	19.32	Marie-Jose Perec, FRA	22.12
2000	Konstantinos Kenteris, GRE	20.09	Marion Jones, USA	21.84
2004	Shawn Crawford, USA	19.79	Veronica Campbell, JAM	22.05

Note to Teachers

Discussion suggestions on the table. *Data are initially presented in this investigation in a table. Students should be directed to reflect on the data, how they were collected, and how they might be used to develop a conjecture related to the initial question of the investigation. To make sense of data, students need to be challenged to explain what* bivariate data *represent and to portray the "story behind the numbers" in this context. The data set provides an opportunity for students to expand their understanding of the organization of data and leads to the development of mathematical equations to estimate future times. The 2008 Olympic times are not included in this table, but they are used at the conclusion of this investigation to examine students' reasoning as they apply the models that they have built.*

As you start the investigation, give students time to reflect individually on how the table is organized and what information is summarized. Then organize the students in small groups to share their thoughts about the information and how they might use it to address the initial question. Finally, groups could share their ideas in a whole-class discussion.

Students' Reasoning about the Table

The following dialogue reflects the reasoning of actual students.

Teacher: The data for the Olympic event are recorded in seconds. Before we discuss the task, explain what data are summarized in this table.

Student 1: The data include the year in which the Olympics was held, the name of the person and country who received the gold medal, and the time for that event.

Teacher: Are there any questions that you have about the data presented in this table?

Student 2: I am interested why were there so many years with only men's times.

Teacher: Women's events were not held in the 200-meter event until 1948. You will also notice that there were years that the Olympics were not held at all. If the 4-year pattern had been followed, there should have been an Olympics in 1916, 1940, and 1944. These games were canceled because of World War I and World War II. The incredible performance of Jesse Owens at the 1936 Olympics is often highlighted as a major event that is connected to World War II—you will notice that he got the gold medal in this event in the last Olympics held before World War II.… In your groups, discuss the question, "Will women ever run faster than men in this race?" How do you suggest that we examine the data represented in the table to answer the question?

Student 1: I think it is actually rather obvious – the women's time for any Olympic year was always slower than the men's time.

Teacher: Yes, but are there any interesting ways in which the relationship is changing?

Student 2: In 1948, the difference in the women's time versus the men's time was 3.3 seconds. In 2004, the difference was 2.26 seconds—women seem to be closing the gap.

Student 3: There are other years, however, where the gap was even larger. Let's calculate this for each of the Olympic years that men and women were involved [*see table 3.3*].

Table 3.3
Gap Times between Men and Women Victors for the Olympic 200-Meter Dash

Year	1948	1952	1956	1960	1964	1968	1972	1976
Difference in times (in seconds)	3.30	3.08	2.80	3.51	2.69	2.75	2.40	2.14

Year	1980	1984	1988	1992	1996	2000	2004
Difference in times (in seconds)	1.84	2.01	1.59	1.80	2.80	1.75	2.26

Student 1: These gaps bounce around, but over time, it seems that the gap is getting smaller.

Students' Reasoning about the Scatter Plot

Table 3.2 shows *bivariate data* collected in time sequence. The table shows measurement observations for two variables, men's running times and women's running times. Bivariate data often can be represented in a scatter plot in the plane, such as the plot in figure 3.1, and lines, or curvilinear functions, can sometimes "fit" the data points. Software tools typically can fit a *least-squares regression line* or a *median-median line* to bivariate data.

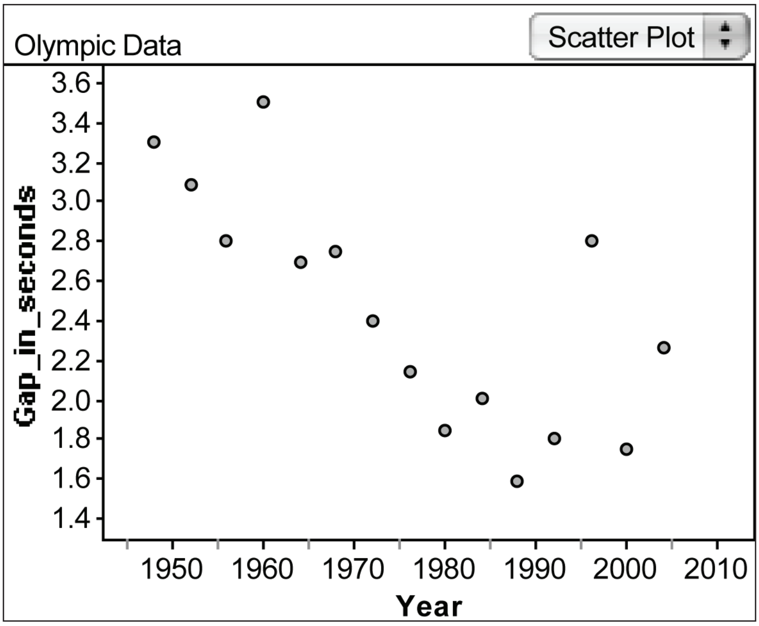

Fig. 3.1. A scatter plot of time differences—gaps—between men's and women's Olympic 200-meter winning times

As the dialogue with the students continues, the least-squares line is superimposed on a scatter plot of the time "gaps" between the men's and women's times plotted over the Olympic years. Students should notice that the points do not really fit the regression line very well and that the distances from the scattered points to the line of fit vary noticeably.

> ***Note to Teachers***
>
> *The source of the actual equations of these least-squares regression or median-median lines is not so important for our investigation. For our purposes, it is more important that students analyze and critique whether or not the line does a reasonable job of fitting the scattered points. In many introductory investigations with data, students can "eyeball" a line to help them develop their intuition about the usefulness of lines to make predictions for what will occur in the future, on the basis of current trends in the data.*

Teacher: We suggested examining the Olympic gaps with a scatter plot [*see fig. 3.1*]. Why do you think a scatter plot might help us answer our questions?

Student 2: A scatter plot shows how the year and the times connect. The year of the Olympics is the x-value, and the difference between the men's and women's times is the y-value.

Student 3: And I think that by organizing the data this way, we can see how the gaps are changing over the years.

Teacher: How are they changing?

Student 1: Although it's not a consistent pattern, the big picture when I stand back looks like the gaps are declining over the years. There are a few years in which the gaps were much wider and threw off this pattern. For example, I wonder why the gaps in 1960 and 1996 were so much greater than the other gaps?

Student 2: Is there a pattern over time that we can use to predict the future?

Teacher: Your question is directly behind what I am asking you to think about. As you mentioned, our x-values of the scatter plot are the Olympic years. Can we use the Olympic years to predict the y-values, or the gaps?

Student 3: I am sure we will not be able to predict it exactly, but maybe we could trace the pattern and come up with a pretty good guess. We would be looking for when the gap is 0 seconds.

Teacher: That is an excellent way to proceed. We'll need to summarize the pattern in some mathematical way.… The computer indicates that the *least-squares regression line* for this scatter plot is

"Differences in time" = (–0.0245) * Year + 50.86.

Look at the line [*see fig. 3.2*]. Let's start with the slope. What is the slope of this line, and what does it mean in our context?

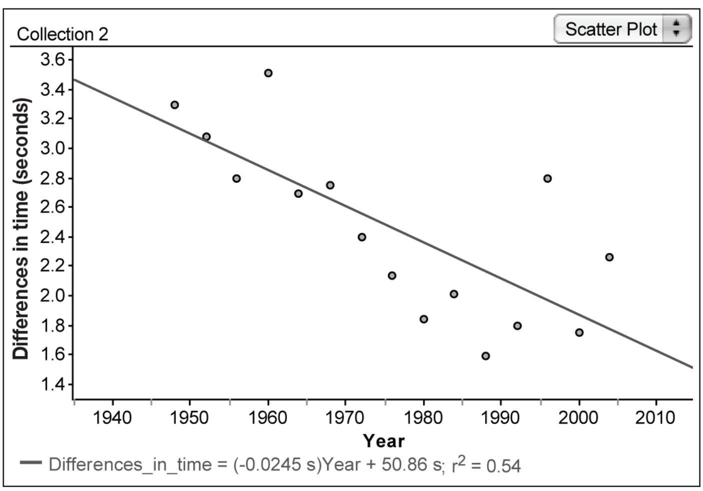

Fig. 3.2. A scatter plot and regression line for differences in men's and women's 200-meter times

Student 1: Well, since this line is going down, the slope is negative. The equation indicates that the slope is –0.0245. The gap is in seconds, and the time is in years. So, the gap between the men and women is going down on average by a little more than 2 hundredths of a second per year.

Student 2: But the Olympics are held only once every 4 years.

Student 1: So if I multiply the slope by 4, I would get an estimate of the change in the gap between Olympics. In that case, the gap looks like it would go down on average by a little more than 8 hundredths of a second. The actual gap times from past Olympics are bouncing around, so I don't know if that is a good indication of what is happening.

Teacher: If the gap were decreasing over time, how could we use it to determine when women would run faster than men?

Student 3: If the gap were 0, then women and men would have approximately the same times. So, if we set the differences in time value in the equation to 0 and then find the year, that would indicate when the men and women would run the race with the same times, or, in other words,

$$\begin{aligned} 0 &= (-0.0245) * \text{Year} + 50.86, \\ -50.86 &= -0.0245 * \text{Year}, \\ -50.86 / -0.0245 &= \text{Year}. \end{aligned}$$

So the year is approximately 2076. This indicates that if the gaps continue to close at the same rate, then in the year 2076—if there is an Olympics in that year—the times should be approximately the same.

Teacher: Do you think the equation for the gap is a good one?

Student 2: No! The points in the scatter plot look like they are going down, but the points don't seem to match the line very well.

Teacher: Very good observation. Let's look at some specific examples. On the one hand, the equation does a good job of predicting the gap in 1964, since the actual gap is 2.69 seconds, and the predicted gap from the equation is 2.74 seconds. That is a difference of 0.05 seconds. On the other hand, it does not do a good job of predicting the gap in 1996. The actual gap for that year is 2.80 seconds, but the predicted gap based on the equation is 1.96 seconds. That is a difference of 0.84 seconds. That seems like a big difference for this event. I wonder why it's so great?

Note to Teachers

Encourage students to reflect on the human factors that are connected with this type of data. For example, could weather or the environment have influenced the times? Urge students to speculate on what other factors might explain the variations in times from the predicted values.

Comparing Plots of Men's and Women's Winning Times in Olympic Years

The next part of the students' conversation is based on their examination of a scatter plot like that in figure 3.3, which displays the actual times for men and women in the 200-meter dash for each of the Olympic years. The scatter plot provides another opportunity to address the investigation's initial question with different regression lines.