

Chapter 1

Introduction

This thesis is concerned with the statistical analysis of sporting performance.

The world over statistics and sport go together. Sport abounds with statistics: baseball and cricket scores, football and basketball scores, and tennis and golf scores are just a few of the statistics that fill up the sports pages of newspapers.

But when measuring sporting performance, sports like these, where players compete to win against each other, cannot compare performances objectively. In each of these sports there have been champions who are claimed to be the best of all time. For example, in baseball Babe Ruth is claimed to be the greatest, in boxing, Muhammad Ali made that claim. Most people would regard Don Bradman as the greatest cricketer of all time, that Rod Laver was the best tennis player, that "Magic" Johnson was the best basketball player, and that "Tiger" Woods is the best golfer ever.

The problem with comparing performers in such sports is that the above sports do not have such standards. To have an objective measure of performance, man must compete against nature, unaided by technology. Athletics and swimming are sports where this happens.

So our goal of comparing sporting performances will use data from swimming and athletics. The main venue for these sports is the Olympic Games.

1.1 The Olympics

The Olympic Games go back a long way.

Though precise records of the ancient Games began only in 776 BC, there is evidence that they existed up to six centuries earlier. Perhaps the first Olympic Games was held at Olympia in 1370 BC. The Olympic Games faded away around the middle of the 9th century BC, but were revived in 776 BC, possibly by King Iphitos of Elis. At that time Greece was a country with many states fighting with each other, and the Olympics was seen to be a way of enforcing a temporary truce among all the warring

factions. A peace pact was proclaimed to last for about three months before the Games (which themselves lasted for five days) and long enough after them for the competitors to enjoy a safe passage back to their homes. The Games of 776 BC, seems to have consisted of merely one event: the stadium race (about 170 meters or 186 yards), won by Coroibos of Olis.

From this time onward, the names and feats of many champions are recorded and competitions in the fine arts were added. The original prizes were only olive wreaths, but gradually the champions began to acquire valuable rewards and the Games became corrupted. The long Roll of Champions ends in 369 AD, and in 393 the Emperor Theodosius decreed from Milan the end of the Olympic Games. Then the Olympic torch went out for 1,503 years.

The idea of review the Ancient Olympic Games was born in Germany. J.C.F. Guts-Muths (1759-1893), the founder of the German gymnastics movement, suggested the idea. However, it was actually Baron Pierre de Coubertin (1863-1937) of France who founded the modern Olympic Games. He was commissioned by the French Government in 1889 to study physical culture throughout the world. His inquiries produced a picture of feuding and dissension between sport and sport, nation and nation, and an already apparent commercial spirit in sport. On 25 November 1892, de Coubertin, in a lecture at the Sorbonne in Paris, for the first time publicly advanced his conviction that there should be a modern revival of the Ancient Games. This led to thirteen countries sending representatives to a conference in 1894 and passing a resolution that "sport competitions should be held every fourth year on the lines of the Greek Olympic Games and every nation should be invited to participate."

A Greek motion was passed giving them the privilege of holding the First Celebration at Athens in 1896. Accordingly, the International Olympic Committee (IOC) then 12 strong - was formed. The small American team won 9 out of the 15 track and field events at the first modern Olympics in Athens in 1896. The Germans dominated the gymnastics, and the French the cycling. The last event - the Marathon - (42.195 km) was dramatically won by Spyros Louis, a post office messenger from Marusi near Athens, one of 21 Greek starters. (McWhirter, 1964).

1.2 Review of Literature

There have been several recent statistical studies of sporting performances. One of the main contributors is Albert (1993, 1994) who studied the career performance of baseballer Mike Schmidt, one of the great home run hitters in major league baseball. Schmidt's hitting performance was notable for its consistency over a span of 14 years. The issue of consistency is investigated by exploring a dataset that lists the date for each homerun Schmidt hit during his career. This exploration suggests that the homerun ability of this hitter did vary within a given season. To see if the observed patterns corresponded to real effects, simulated homerun data was produced assuming a coin tossing model with a constant probability of success. By comparison of Schmidt's data with this simulation distribution, it is seen that Schmidt's homerun hitting behaviour did exhibit some inconsistency. These patterns show how Schmidt matured as a baseball hitter over his career.

Albright (1993) and Stern and Morrid (1993) investigated baseball hitting streaks. In fact baseball, the most popular sport in the United States, has attracted quite a lot of interest from statisticians. A long time ago, Lindsey (1963) used mathematical models to compare baseball strategies. Some more recent references include James (1982), Bennett & Flueck (1984), Barry & Hartigan (1993), Thorn & Palmer (1993), Cassella & Berger (1993), and Bennett (1993).

Berry & Larkey (1998) recently investigated the effects of age on the performance of professional golfers, and concluded, not surprisingly, that 'in golf, it is very clear that age will overtake your ability.

Turning to the comparison of performances over different periods of time, the objective of our investigation, not many studies has been done. Lunn & McNeil (1991) compared Olympic men's sprinting performances from 1900 to 1988. More recently, Berry, Reese and Larkey (1999) developed a method for bridging different eras in sports.

1.3 Scope of Investigation and Research Objectives

Our source of data is the set of winning results in all individual men's and women's swimming events, and in most individual men's and women's running, jumping and throwing events, at Olympic Games from 1928 to 2000.

We have excluded long-distance running events (greater than 1500 meters), and the pole vault event (because its results are unduly affected by technology).

Our main objective is to create a model of each sport that fits, and thus forecasts the Olympic winner results in swimming, running, jumping and throwing.

Our secondary objective is to use these models of four sports to compare the performances in the different sports and between men and women over the years.

In Chapter 2 we describe the analysis methods.

Chapter 3 is concerned with preliminary analysis of the data, while the model combining the results is developed in Chapter 4.

Finally, Chapter 5 gives the main conclusions, and some limitations of our study.