

STAT 503 Assignment 1: Tennis Statistics

Due: Sketch of report, with plan and summary statistics in class Feb 2
Full report due Feb 7.

The Australian Open Tennis tournament has been playing in Melbourne for the last week. The grand finals are this weekend. The web site <http://www.australianopen.com> contains a lot of information and statistics about the players in the tournament.



There have been some interesting developments during the tournament. Serena Williams (USA) is having a serious comeback. After winning several grand slams a few years ago, she's been off the circuit with injury and family tragedy in the last two years. She came into this tournament as a wild card, and her international ranking has grown from the 80s into the top 20s. She plays Maria Sharapova (RUS) on Friday night. On the mens side, Roger Federer (SUI) has dominated tennis for four years now, and many say that he is the best player of all time, that he's a tennis genius. His play is fabulous to watch and devastating for opponents, as he carefully dissects their game. He will play the grand final on Sunday morning (2:30am) against a lesser known player from Chile, Fernando Gonzalez. Gonzalez has changed coaches in the last 6 months and has streaked through this tournament with play that mirrors Federer in strokes and intelligence. He made short matches defeating the top seeds in his side of the draw.

The data for this assignment comes from the Australian Open web site. The data is organized into two files for womens and mens results. Each row contains statistics for one player, which are numbers for all the matches the player has played in the tournament averaged. Each player played in at least three matches, that is, survived to Round 3.

Your main job is to answer this question:

“Which player statistics best predict of performance in the tournament?”

This assignment requires you to construct your own report on this data with the members of your working group. Specifically, I need you to *write down two or more questions* that might be addressed by this data. You should try to do this ahead of time before, doing any plots or summaries. You should search for additional information for yourselves on the web: eg what's your prediction for the winner of the grand finals.

You should follow the format used for Case Study 1 (Restaurant Tipping) for your report. You will need to present your results in tabular and graphical form, and concisely summarize the major findings. Please just use simple methods for this assignment, no more than regression, even though some multivariate methods might be appropriate you'll get opportunities to use these in the next assignment.

In keeping with doing reproducible research, you should also submit your R code and modified data to the assignment page in WebCT, so that I can re-do your analysis. Your report needs to be handed in on paper, and emailed to me.

The grade for the assignment will depend on neatness of the report, how comprehensive is the analysis (are your questions answered), clarity of results presentation, and preciseness of final conclusions.