

What the numbers say:

Can we use statistical analysis and probability theory to predict the winning times for the USA Olympic Swimming Trials in July? Although it's impossible to know for sure the winning time (until the races have actually taken place), it is possible to look back in history and use statistics to predict what it *might* take to win. It's no surprise that competitive swim times have been getting progressively faster. Performances should improve, as better nutrition, better coaching, and faster pools all contribute to a swimmer's success!

Consider the graph below of the men's 100 meter freestyle. The average time for the eight finalists in the Men's Olympic trials from 1968 to 2000 have been plotted. Clearly the *average time* is coming down, but not as much in recent years as it did in the past. The best way to fit a line through these points is a 'power curve' (as compared to a straight line). In this case the equation described by the data is: $\text{Time} = 53.681 * \text{Year}^{-0.0366}$. How closely data points



follow a curve is described with the correlation coefficient "R". When R is squared it allows for an estimation of the variance explained by the relationship. In this case R^2 is 0.97, or almost a perfect fit (perfect is 1.00). The conclusion then is that the line described by the equation fits the points very well. Extending the curve out to the 2004 Trials allows us to predict that the average of the top eight times in the men's 100 free will be 49.34 " 0.25 seconds. We need the " 0.25 second range (standard deviation) to account for chance, but we are 95% confident that the interval from 49.09 to 49.59 will contain the *average of the top eight times* for this event.

But what will it take to *win*? Since there's only one winning time per year, this doesn't give us many numbers to work with, and thus, our ability to predict the winning time is not very good. Another approach then, is to calculate how much faster *the winner* has been than *the average* of the top 8 in the past. This is expressed as a Z score, or the number of standard deviations the winner has been from the average. The average Z score for the winner of the men's 100 free from 1968 to 2000 was -1.50, which corresponds to a predicted winning time of

48.51 for the 2004 Trials. We still need to express this as a range (probability again), so we say that we are 95% confident that the interval 48.26 to 48.76 will contain the winning time. The same calculations have been done for all of the other events and are presented in the table below.

But wait just a minute! Some of our potential winners have already gone faster than many of the predicted winning times. That's great because with the information provided in the table and the names of these athletes, we can now predict which swimmers might be the 'odds on favorite' to qualify for the Olympics. Keep in mind that the method we are using to predict the 2004 Trials does not account for anything that has happened in swimming since 2000. For instance, rule changes would have a significant effect upon this process. In addition, we have seen the rise of some phenomenal performers, for example, Michael Phelps and Natalie Coughlin. As far as statistics go, the performances of these swimmers represent *outliers*, or more than the average Z score below the average times of the past. This is true for all record performances and outstanding performers. Statistics cannot predict very well what they will do. But, it will certainly be fun to watch!

Predicted Winning Times for the 2004 US Olympic Trials Swimming Events										
	Predicted		Winning	95% Conf. Int.		American	AR		World	Olympic
Men's Events	Mean Top 8	" S.D.	Time	Lower	Upper	Record	Holder	Year	Record	Record
50 Free	:22.15	0.13	:21.73	:21.60	:21.85	:21.76	Hall	2000	:21.64	:21.91
100 Free	:49.34	0.25	:48.51	:48.25	:48.76	:48.33	Ervin	2001	:47.84	:47.84
200 Free	1:48.08	0.67	1:46.54	1:45.87	1:47.21	1:45.99	Phelps	2003	1:44.06	1:45.35
400 Free	3:53.02	0.99	3:48.94	3:47.95	3:49.93	3:46.73	Phelps	2003	3:40.08	3:40.39
1500 Free	15:22.41	2.93	15:05.35	15:02.42	15:08.28	14:56.81	Thompson	2000	14:34.56	14:43.48
100 Back	:55.16	0.24	:53.94	:53.70	:54.18	:53.60	Krayzelburg	1999*	:53.60	:53.72
200 Back	1:59.64	0.90	1:57.29	1:56.39	1:58.19	1:55.15	Peirsol	2002*	1:55.15	1:56.76
100 Breast	1:01.91	0.15	1:00.70	1:00.55	1:00.86	1:00.21	Hansen	2003	:59.78	1:00.46
200 Breast	2:14.22	0.70	2:11.71	2:11.01	2:12.41	2:10.16	Barrowman	1992	2:09.42	2:10.16
100 Fly	:53.46	0.25	:52.50	:52.24	:52.75	:50.98	Crocker	2003*	:50.98	51.96
200 Fly	1:58.95	0.15	1:56.78	1:56.63	1:56.93	1:53.93	Phelps	2003*	1:53.93	1:55.35
200 IM	2:02.52	0.66	2:00.11	1:59.45	2:00.77	1:55.99	Phelps	2003*	1:55.94	1:58.98
400 IM	4:20.41	0.80	4:13.86	4:13.06	4:14.65	4:09.09	Phelps	2003*	4:09.09	4:11.76
	Predicted		Winning	95% Conf. Int.		American	AR		World	Olympic
Women's Events	Mean Top 8	" S.D.	Time	Lower	Upper	Record	Holder	Year	Record	Record
50 Free	:25.58	0.16	:25.01	:24.86	:25.17	:24.63	Torres	2000	:24.13	:24.13
100 Free	:55.10	0.24	:54.22	:53.98	:54.47	:53.99	Coughlin	2002	:53.77	:53.77
200 Free	2:01.26	0.11	1:59.69	1:59.59	1:59.80	1:57.41	Benko	2003	1:56.64	1:57.65
400 Free	4:14.36	1.35	4:09.85	4:08.50	4:11.20	4:03.85	Evans	1988*	4:03.85	4:03.85
800 Free	8:38.80	2.13	8:28.76	8:26.63	8:30.89	8:16.22	Evans	1989*	8:16.22	8:19.67
100 Back	1:02.29	0.13	1:01.36	1:01.23	1:01.49	:59.58	Coughlin	2002*	:59.58	1:00.21
200 Back	2:15.05	0.45	2:11.71	2:11.26	2:12.15	2:08.53	Coughlin	2002	2:06.62	2:07.06
100 Breast	1:09.09	0.43	1:07.18	1:06.74	1:07.61	1:07.05	Quann	2000	1:06.37	1:07.02
200 Breast	2:28.30	0.40	2:24.38	2:23.97	2:24.78	2:22.99	Beard	2003*	2:22.99	2:24.35
100 Fly	:59.62	0.21	:58.20	:57.99	:58.42	:57.58	Torres	2000	:56.61	:56.61
200 Fly	2:13.19	0.73	2:09.37	2:08.65	2:10.10	2:05.88	Hymann	2000	2:05.78	2:05.88
200 IM	2:15.97	0.13	2:12.64	2:12.51	2:12.77	2:11.91	Sanders	1992	2:09.72	2:10.68
400 IM	4:48.14	1.56	4:39.11	4:37.54	4:40.67	4:37.58	Sanders	1992	4:33.59	4:33.59

SD = Standard Deviation 95% Conf. Int. = 95% Confidence Interval

AR = American Record * Indicates American Record is World Record