
AWESOME BABY!!!


WWF….SMACKDOWN!!!!!
AWESOME BABY!!!!


Duke over UCLA by 3! Temple over UCONN by 7!
She still thinks I’m a freakin’ Husky!!!!!

March Madness And The Office Pool

Edward H. Kaplan

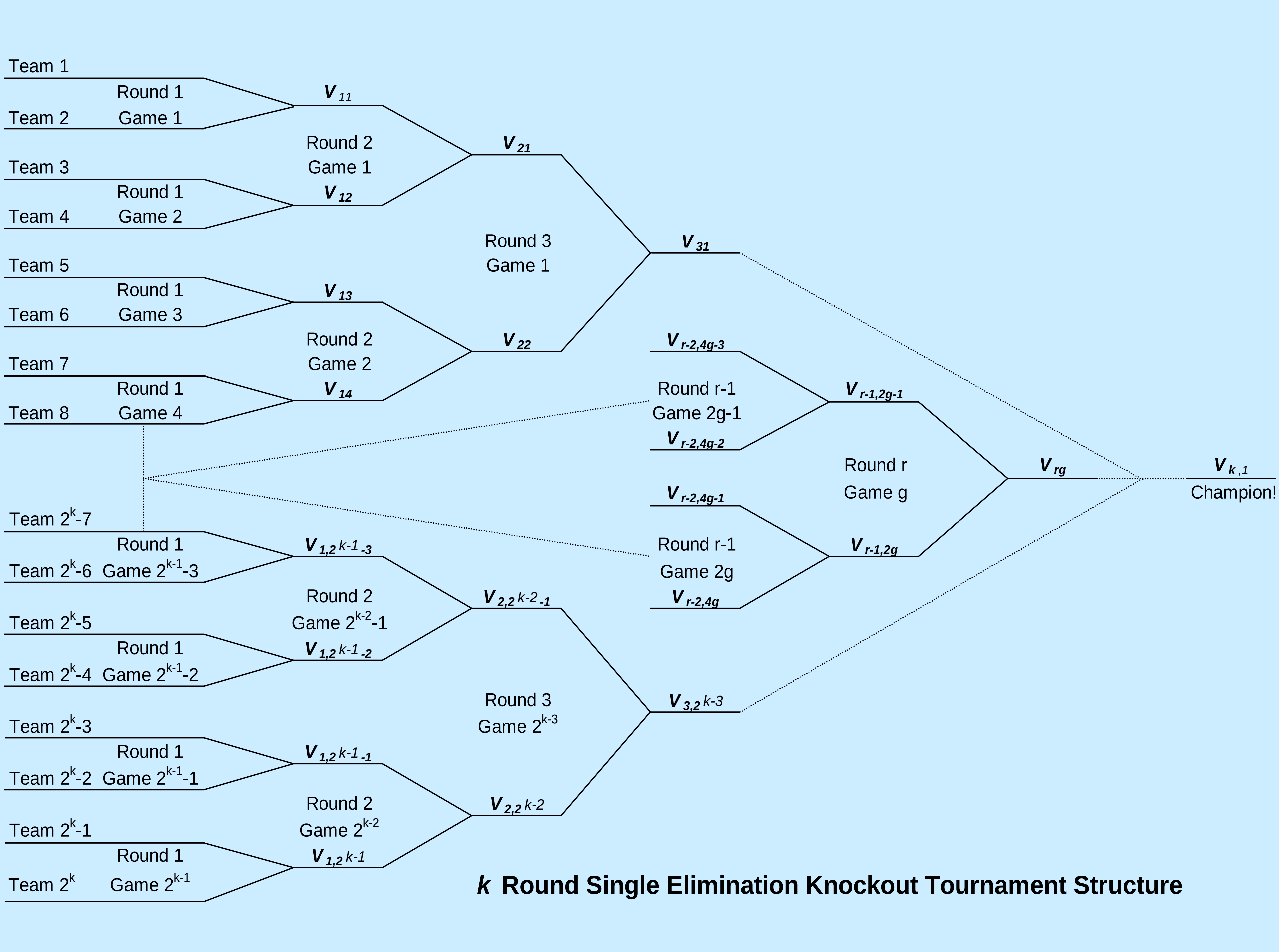
Stanley J. Garstka

Yale University

School of Management

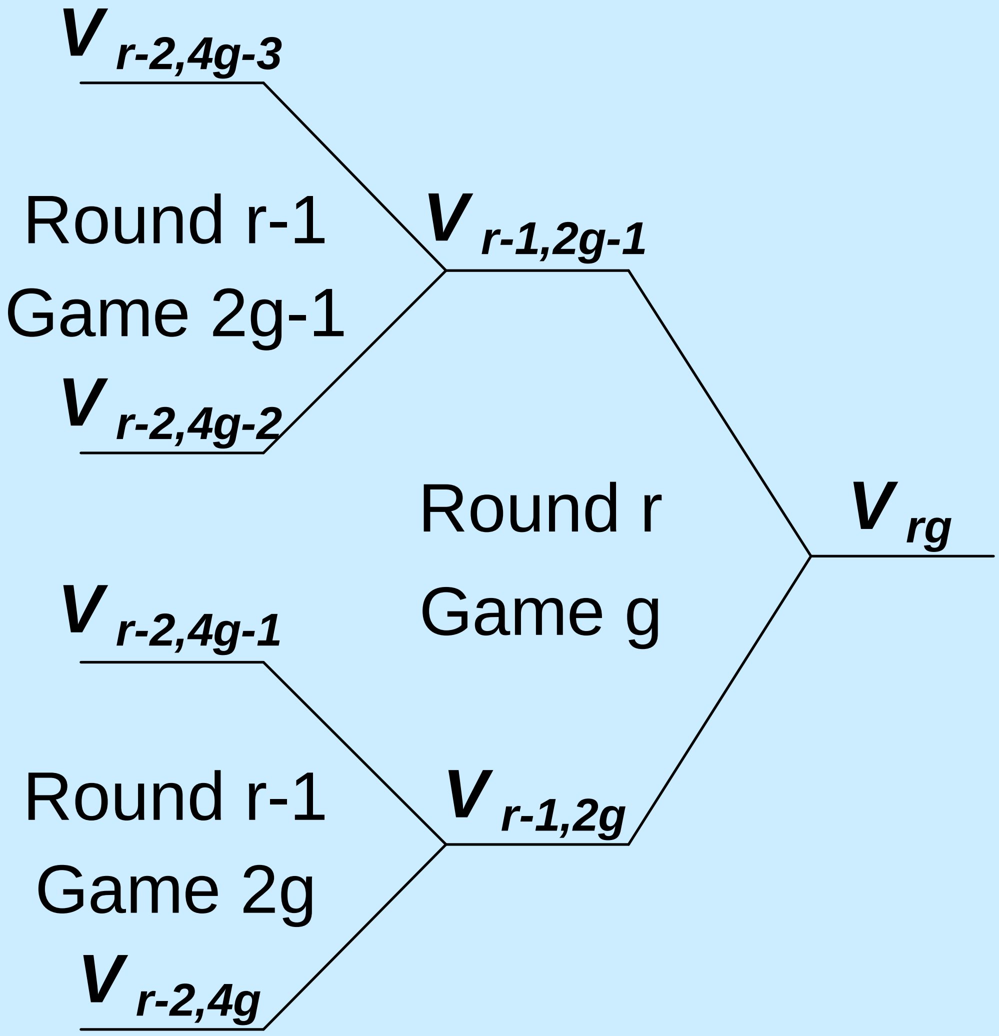
What We're Going To Do:

- Describe tournament structure
- Embed office pools in tournament structure
- Illustrate with a random tournament (let Alli Baby pick the winners!)
- Illustrate with toy Markov tournaments
- Show how to maximize the expected number of points in a real pool
- Markov models for college tournaments
- Empirical results



***k* Round Single Elimination Knockout Tournament Structure**

Single Round Elimination Tournament Structure



The Victor!

Tournament Structure Facts

- Tournaments with k rounds start with 2^k teams, and require 2^k-1 games
- e.g. NCAA tournament has 6 rounds, 64 teams, and 63 games
- The number of possible realizations equals 2^{2^k-1}, so for the NCAA tournament there are 2^{63} or about 9.22×10^{18} (that is, more than 9 billion billion) possible realizations

Office Pools

- Points are awarded for correctly naming game winners *a priori*
- Score for the pool is the sum of the points awarded for each game
- (S)he who gets the most points wins
- Simplest case: pick as many games correctly as possible
- More complex case: points increase with tournament depth; upset points awarded

Example: A Random Tournament

Example: A Random Tournament

- For a team to win in round *r*, it must win *r* games in a row! This has probability $1/2^r$
- In round *r* of a *k* round tournament, there are 2^{k-r} games
- The expected number of games called correctly in round *r* is thus $2^k/4^r$
- The expected fraction of correct calls goes to 1/3 (geometric series in 1/4)!
- Variance (# correct) goes to mean / 1.05

Possible Objectives

- Maximize expected winnings (money is everything)
  - Need to worry about other players

Possible Objectives

- Maximize expected winnings (money is everything)
  - Need to worry about other players
- Maximize expected number of points in the pool (we want to be right!)
  - Does not depend on the behavior of others
  - Does depend on probability assessments for game outcomes, and pool rules for points

Example: 2001 Euro Final Four

Row Beats Column With Probability:				
	--	0.6	0.6	1
	0.4	--	0.6	1
	0.4	0.4	--	1
	0	0	0	--

- Draw: Πανα/Pilsen, מכבי/ЦСКА, winners
- Who should you pick to maximize the expected number of wins?

Example: 2001 Euro Final Four

Row Beats Column With Probability:				
	--	0.6	0.6	1
	0.4	--	0.6	1
	0.4	0.4	--	1
	0	0	0	--

- Clearly choose מכבי over ЦСКА
- Πανα or Pilsen?
- Who should you pick for the final?

Example: 2001 Euro Final Four

- To maximize expected number of correct picks, pick Παvα over Pilsen, מכבי over ЦСКА, and מכבי over Παvα (expected wins: .6 + 1 +.4 = 2)
- Optimal to pick מכבי over Παvα, *even though* Παvα *beats* מכבי *w.p. 0.6!*
- Really picking מכבי to win final, *not* מכבי to beat Παvα!
- Picking Παvα, מכבי, Παvα yields an expected .6 + 1 + .6 × .6 = 1.96 wins

Example: Impact of Depth Points

Row Team Beats Column Team With Probability:

	A	B	C	D
A	--	0.5	0.6	0.2
B	0.5	--	0.7	0.2
C	0.4	0.3	--	0.6
D	0.8	0.8	0.4	--

- A plays B, C plays D, winners play
- What picks maximize E(# wins)?
- What picks maximize Pr{Pick WINNER!}?

Example: Impact of Depth Points

Row Team Beats Column Team With Probability:

	A	B	C	D
A	--	0.5	0.6	0.2
B	0.5	--	0.7	0.2
C	0.4	0.3	--	0.6
D	0.8	0.8	0.4	--

- To maximize E(# wins), choose B-C-B (yields expectation of 1.35)
- Team D has highest chance (32%) of winning (pick A-D-D or B-D-D)

Example: Impact of Depth Points

Row Team Beats Column Team With Probability:

	A	B	C	D
A	--	0.5	0.6	0.2
B	0.5	--	0.7	0.2
C	0.4	0.3	--	0.6
D	0.8	0.8	0.4	--

- Suppose you get 1 point for calling each of first two games, *d* points for championship
- If $d < 20/7$, pick B-C-B, otherwise pick A-D-D or B-D-D

Optimizing in Rea

- Let $\mu_{rg}(i)$ denote the ex points through game g team i to win that gam
- Let μ^*_{rg} denote the exp optimally thru game g
- Let $\pi_{rg}(i)$ be the points correctly picking team round r
- Let $\omega_{rg}(i) = \Pr\{i$ wins

Optimizing in Rea

- Let $\mu_{rg}(i)$ denote the e
points through game g
team i to win that gam
- Let μ^*_{rg} denote the exp
optimally thru game g
- Let $\pi_{rg}(i)$ be the points
correctly picking team
round r
- Let $\omega_{rg}(i) = \Pr\{i$ wins

Optimizing in Real Office Pools

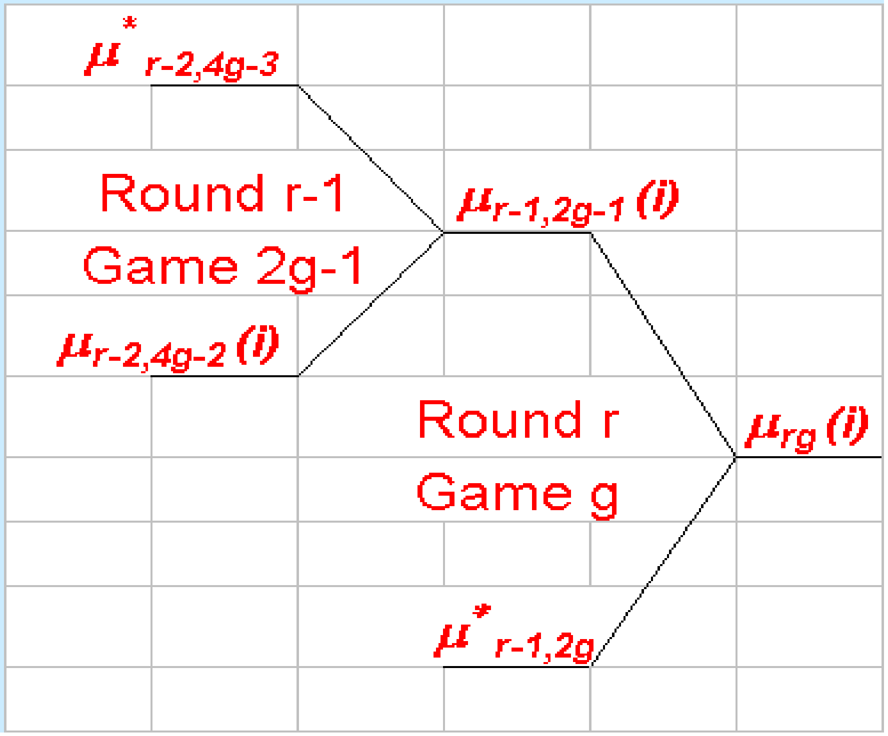
- Let $\mu_{rg}(i)$ denote the expected number of points through game g in round r if choose team i to win that game and act optimally
- Let μ^*_{rg} denote the expected points gained optimally thru game g in round r
- Let $\pi_{rg}(i)$ be the points awarded for correctly picking team i to win game g, round r
- Let $\omega_{rg}(i) = \Pr\{i \text{ wins game } g, \text{ round } r\}$

Optimizing in Real Office Pools

$$\mu_{rg}(i) = \mu_{r-1,2g-1}(i) + \mu^*_{r-1,2g} + \pi_{rg}(i)\omega_{rg}(i) \quad \text{if } i \text{ plays in } r-1, 2g-1$$

$$\mu_{rg}(i) = \mu^*_{r-1,2g-1} + \mu_{r-1,2g}(i) + \pi_{rg}(i)\omega_{rg}(i) \quad \text{if } i \text{ plays in } r-1, 2g$$

$$\mu^*_{rg} = \max_{i \in \tau(r,g)} \mu_{rg}(i)$$



Markov Models for College Basketball

- We need to compute the probabilities that any team i wins game g in round r ($\omega_{rg}(i)$)
- We do this via Markov models
- Markov assumption: in any game between two teams i and j, the probability that i beats j equals p_{ij} independently of other games

$$\omega_{rg}(i) = \omega_{r-1,2g-1}(i) \sum_{\ell \in \phi(i,r)} \omega_{r-1,2g}(\ell)\, p_{i\ell}$$

Markov Models for College Basketball

- We consider three models:
  - regular season record
  - expert ratings (Sagarin ratings)
  - Las Vegas Odds

Regular Season Model

- $p_{ij} = s_i / (s_i + s_j)$ (Bradley-Terry model, or model of *quasi-independence*)
- Estimate the parameters via (constrained) maximum likelihood based on regular season (and conference tournament) results
- Use 64 NCAA teams, 32 NIT teams, and all others lumped into *megateam* (so 96 free parameters)
- Provides connectivity among all teams

3 Team Example

- *a* plays *b*, *c* plays *b*, but *a* and *c* never play
- f_{ab} and f_{cb} are respectively the observed fraction of time *a* and *c* beat *b*
- Solve 3 equations:

$$\frac{s_a}{s_a + s_b} = f_{ab}, \frac{s_c}{s_b + s_c} = f_{cb}, s_a + s_b + s_c = 1$$

- Implied s_a depends on performance of *c*!

$$s_a = \frac{f_{ab}(1 - f_{cb})}{1 - f_{ab} f_{cb}}$$

Sagarin Ratings

- Sagarin provides estimates of scoring rates λ_i for all Division I NCAA teams
- We assume *uncorrelated Poisson scoring!*
- Let X_{ij} be the point spread. Under the model, X_{ij} is approximately normal with mean λ_i-λ_j and variance $\lambda_i+\lambda_j$
- p_{ij}= Pr$\{X_{ij} > 0\}$

Las Vegas Odds

- Las Vegas takes bets on NCAA games
- Two types of bets: point spreads and point totals ("over/under" bets)
- If market is correct, then can directly estimate scoring rates from quoted point spreads and point totals, and use uncorrelated Poisson model
- Implies that the actual point spread that will occur is normally distributed with mean given by Vegas point spread line, and variance given by Vegas over/under line!

Tournament by Tournament

	NCAA 1999	NCAA 1998	NIT 1999	NIT 1998
	Actual/Expected/Standard Deviation of Wins	Actual/Expected/Standard Deviation of Wins	Actual/Expected/Standard Deviation of Wins	Actual/Expected/Standard Deviation of Wins
Chance	-- / 21.3 / 4.5	-- / 21.3 / 4.5	-- / 10.7 / 3.2	-- / 10.7 / 3.2
Pick the Seeds	36 / -- / --	39 / -- / --	17 / -- / --	13 / -- / --
Regular Season	39 / 42.6 / 4.3	37 / 43.5 / 4.1	17 / 12.9 / 3.4	18 / 15.0 / 3.4
Sagarin	41 / 41.4 / 4.2	39 / 38.5 / 4.5	15 / 13.1 / 3.4	13 / 13.0 / 3.4
Las Vegas Odds	38 / 44.6 / 3.8	35 / 45.1 / 4.3	22 / 17.5 / 3.7	15 / 19.2 / 3.7

Packard Pool

- Run by a math professor at Mesa State U.
- Points awarded if correctly call n^{th} seed to win in round r given by $(1+n)\rho(r)/2$ where $\rho(r)$ increases nonlinearly as r goes from 1 thru 6 ($\rho(1) = 1{,}890$ while $\rho(6) = 27{,}720$)
- How would we have done?

The Packard Pool!

	NCAA 1999	NCAA 1998
	Actual/Expected/Standard Deviation of Total Score	Actual/Expected/Standard Deviation of Total Score
Pick the Seeds	239,256 / -- / --	233,702 / -- / --
Regular Season	274,956 / 302,964 / 47,293	356,969 / 304,181 / 54,213
Sagarin	405,421 / 299,062 / 55,738	274,572 / 273,333 / 63,974
Las Vegas Odds	299,942 / 339,887 / 62,735	213,285 / 514,516 / 138,543

- Winner of 1999 pool scored 333,572 points
- Our Sagarin model would have won!
- Winner of 1998 pool scored 218,545 points
- Our Regular Season and Sagarin models would have won (as would picking the seeds)!

2000 NCAA Tournament Results

		CBS Sportsline Pool	Packard #2 Pool	ESPN Pool
# Participants		95,000	40	580,000
Upset Points?		Yes	Yes	No
	Percentile	99.97	92.5	71.3
Massey Ratings	x / μ / σ	426 / 259 / 76	341,807 / 275,271 / 59,876	780 / 799 / 203
	# Wins	39	38	37
	Percentile	83.1	90	61.9
Sagarin Ratings	x / μ / σ	228 / 241 / 105	304,304 / 268,067 / 66,167	710 / 738 / 193
	# Wins	29	38	39
	Percentile	40.2	70	32.5
Regular Season	x / μ / σ	173 / 248 / 93	262,872 / 277,201 / 76,480	530 / 713 / 185
	# Wins	37	31	
	Percentile	95.8	80	28.4
Las Vegas Odds	x / μ / σ	263 / 258 / 54	281,388 / 315,024 / 51,254	510 / 1,027 / 196
	# Wins	40	40	39

So, Did We Make Any Money?

UNIVERSITY OF
CONNECTICUT
STORRS, CONNECTICUT 06269

51-44
119
FLEET BANK
ACCOUNTS PAYABLE ACCOUNT

CHECK NO. 15-645073

CHECK DATE	CHECK AMOUNT
03/26/00	*****$250.00

VOID AFTER 120 DAYS

PAY *Two hundred fifty and 00/100 Dollars*

TO THE ORDER OF
EDWARD H KAPLAN
YALE SCHOOL OF MANAGEMENT
PO BOX 208200
NEW HAVEN CT 06520-8200

SECOND SIGNATURE REQUIRED ABOVE $9,999.99

AUTHORIZED SIGNATURE

THE BACK OF THIS DOCUMENT CONTAINS AN ARTIFICIAL WATERMARK - HOLD AT AN ANGLE TO VIEW

⑈645073⑈ ⑆011900445⑆ 6772 8052⑈

How About 2001?

- Tried to predict *women's* tournament
- After first two rounds, the model correctly picked 42 of 48 games, beating 99.7% of 60,831 entries in *ESPN* pool
- But then…
- Of remaining 15 games, model got only 4 correct, ending up at the 48.3%-ile

2001 Men's Tournament

- Highlights: beat 71.3% in the 600,000+ *ESPN* pool
- Came in 3rd place!!!…
- …in the Yale School of Management pool, winning $17.50
- And, Stan and I won *four T-shirts* in the Fantasy Cup Tournament…
- …for coming in 180th out of 1200+
- The winner got $5,000

Web Implementation

- Tom Adams' "poologic" calculator at

  http://www.poologic.com/

SO.....

Wanna dance??????