America's Idol?
How the Contestant Most Voted for Doesn't Always Win

By: Jason Gershman

Assistant Professor of Mathematics and Statistics and Coordinator of Mathematics

Nova Southeastern University

Outline

- How the Show Works
- The Data
  - History of the Data
  - How it's Collected
  - How it's Been Verified
  - Flaws and Corrections
- Unfairness
  - Geography
  - Performance Order

American Idol

- Spinoff of Britain's Pop Idol
- Season 9 aired from January – May 2010 on Fox
- The Highest Rated Show for 5 straight seasons

Structure of American Idol

- The final 12 contestants are eliminated one at a time (one per week) until the winner is crowned
  - contestants sing during Tuesday's show ("Performance Show")
  - afterwards, the American public votes via toll-free phone lines or text messaging for two hours
  - results are announced during Wednesday's show ("Results Show")
  - contestant with the lowest number of votes is eliminated

Important Notes

- Vote totals are not released to the public
  - Vote tallies are certified by a contracted accounting firm.

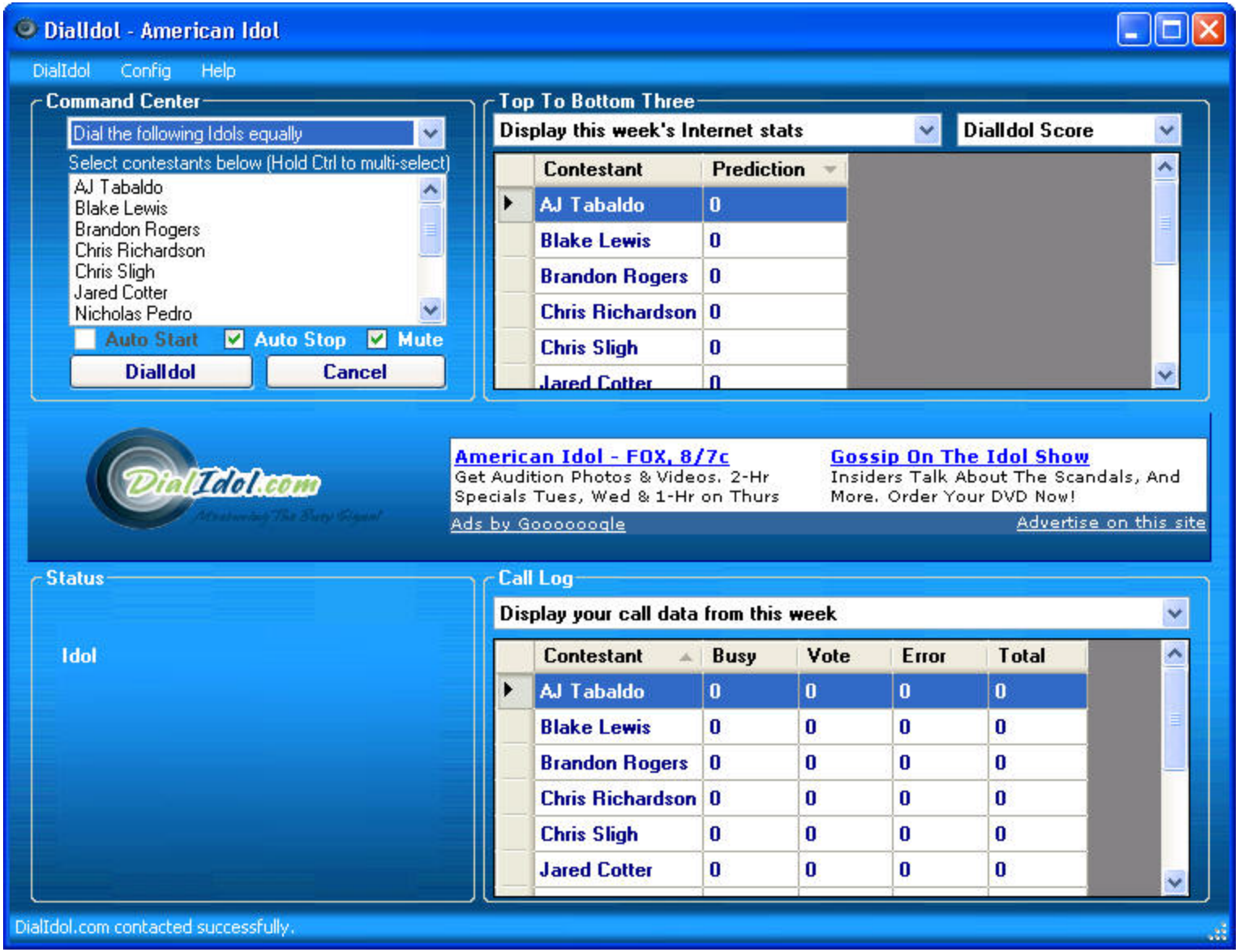
- You can't vote for whom you want to eliminate only for whom you want to keep
  - Polarizing Contestants last longer than those who are labeled "forgettable" by judge Simon Cowell

DialIdol

- Created by computer programmer Jim Hellriegel
  - Works as a modem dialer to give fans an easy way to vote
  - Computer's telephone modem repeatedly dials your favorite contestant's phone number
  - Created during Season 4 in the Winter of 2005

- Not a "power dialer" which would be illegal according to the American Idol rules

DialIdol


DialIdol - American Idol
DialIdol
Config
Help
Command Center
Dial the following Idols equally
Select contestants below (Hold Ctrl to multi-select)
AJ Tabaldo
Blake Lewis
Brandon Rogers
Chris Richardson
Chris Sligh
Jared Cotter
Nicholas Pedro
Auto Start
Auto Stop
Mute
DialIdol
Cancel
Top To Bottom Three
Display this week's Internet stats
DialIdol Score
Contestant
Prediction
AJ Tabaldo 0
Blake Lewis 0
Brandon Rogers 0
Chris Richardson 0
Chris Sligh 0
Jared Cotter 0
DialIdol.com
American Idol - FOX, 8/7c
Get Audition Photos & Videos. 2-Hr Specials Tues, Wed & 1-Hr on Thurs
Gossip On The Idol Show
Insiders Talk About The Scandals, And More. Order Your DVD Now!
Ads by Gooooooogle
Advertise on this site
Status
Idol
Call Log
Display your call data from this week
Contestant
Busy
Vote
Error
Total
AJ Tabaldo 0 0 0 0
Blake Lewis 0 0 0 0
Brandon Rogers 0 0 0 0
Chris Richardson 0 0 0 0
Chris Sligh 0 0 0 0
Jared Cotter 0 0 0 0
DialIdol.com contacted successfully.

What DialIdol Does

- DialIdol's servers receive the following information for each call made:
  - which idol's phone line your computer called
  - whether the phone line was busy or the vote was received by American Idol

What Does DialIdol Report?

- DialIdol began publishing the “results”
  - number of votes for each idol
  - number of busy signals received
  - Results were originally published to allow fans to see which Idols were “in trouble” of elimination to have their fans vote for them more.
  - Jim began using these results to make predictions on who would be eliminated.

Success and Controversy

- Dialidol began using statistics to predict who would be eliminated
  - Accurate about 65% of the time during Season 4 of American Idol
  - Sample Sizes were small (the website was not well known)

- Early in season five, DialIdol was being sued by Fox for copyright infringement and Jim received a Cease and Desist Letter

What Happened

- Fox felt the website was ruining the element of surprise
  - Fox was afraid that the website would be a spoiler and tell the world who would be eliminated before Ryan Seacrest had a chance to on Wednesday's show
  - They feared losing their audience and felt threatened by the increasing traffic at the DialIdol Website

The Aftermath

- After consultation with his lawyers, Mr. Hellriegel was able to keep Dialidol going after making some superficial changes to his website to avoid copyright infringement
  - Media attention increased hits at Dialidol
    - Jim was on MSNBC, CBS, Fox News, CNN
    - DialIdol was a bookmarked must for all idol fans
  - The number of votes made using Dialidol rose by 2000% in one month in the middle of Season 5 in 2006.
  - Larger sample size = Better predictions

Dialidol's Big Week

- Any doubts of the website's accuracy was erased by the results from the "final four" week in season 5
  - 4 idols remained: Chris Daughtry, Taylor Hicks, Elliott Yamin, and Katharine McPhee.
  - In a "shocker," Chris Daughtry was eliminated and sent home.
  - Shocking result if you did not analyze the data; Not shocking to those who did analyze the data

The Data

Idol	Busy Signals	Votes	Totals
Contestant 1	63406	22960	86366
Contestant 2	53325	32300	85625
Contestant 3	52752	14866	67618
Contestant 4	49501	30915	80416

Who do you think got eliminated this week?

Think of some ways to properly use this sample data to make meaningful inference on the population

Sample Biases

- Sample is biased:
  - towards those who care enough to use the site to vote repeatedly
  - towards those who own a computer with a modem dialer
  - against those who vote via cell phone
  - against those who vote via text message
- The Top 3 biases are not a problem…the last one is an increasingly bigger problem each season.

Making Predictions

- Need to understand the role of the busy signal
  - Is a busy signal a “good” or a “bad” thing?
- A busy signal is “good” for that idol
  - It means other people are voting for him…if you can’t get through, everyone else is calling for him…if you can get through every time, it means he has no other fans voting for him except you.
- This result avoids bias in the sample
  - Measures the population results without dependence on the sample results
  - It doesn’t matter who you vote for but it acts as a proxy of how busy a contestant’s phone line is.

The “Shocker”

Idol	Busy Signals	Votes	Totals	Busy Percentage
Contestant 1	63406	22960	86366	73.4
Contestant 2	53325	32300	85625	62.3
Contestant 3	52752	14866	67618	78.0
Contestant 4	49501	30915	80416	61.6

The Idols were:

Idol 1= Elliott

Idol 2= Katharine

Idol 3= Taylor

Idol 4= Chris, who was eliminated

Season 5: Championship Week

Idol	Busy Signals	Votes	Totals
Contestant 1	185342	99497	284839
Contestant 2	361451	78939	440390

Who won American Idol?

And the Winner Is…

- Contestant 1, Katharine McPhee, had a 65.1% busy percentage
- Contestant 2, Taylor Hicks, had an 82.1% busy percentage
- Taylor Hicks won American Idol in a landslide

What Has Changed

- Dialidol was 87% accurate in Season 5, 88% accurate in Season 6 but has declined in accuracy in the past two seasons due to a rise in text messaging, which still cannot be accounted for

- Fox, realizing it cannot stop Dialidol, sometimes plays with the data of the “bottom three” or “bottom two” to try to show Dialidol is not perfect but it has not hurt the ratings…in fact, ratings have gone up

Role of Predicting Games/Markets

- Part of the popularity with Dialidol, is it's use in gaming
  - Offshore Prediction Markets like Intrade use the data from Dialidol to set their betting lines for customers wishing to wager on the outcome of the show
  - Players in high stakes office pools use the data from Dialidol to win money from legal office pools.

Geography

- One element of unfairness in American Idol is geographic bias

- The audience has two hours to vote at the conclusion of the show when it has aired in their time zone

Timeline (all in EDT)

- From 10:00-11:00 PM, the Eastern and Central Time Zones (78% of the country) can vote
- From 11:00 PM-12:00 AM, the Eastern, Central, and Mountain Time Zones (84% of the country) can vote
- From 12:00 AM-1:00 AM, the Mountain Time Zone (6% of the country) can vote)
- From 1:00-3:00 AM, the Pacific and Alaskan Time Zones (15.5% of the country) can vote
- From 3:00-5:00 AM, the Hawaiian Time Zone (0.5% of the country) can vote
- Note: Puerto Rico and the US Virgin Islands watch the show live and vote with the Eastern and Central Time Zones.

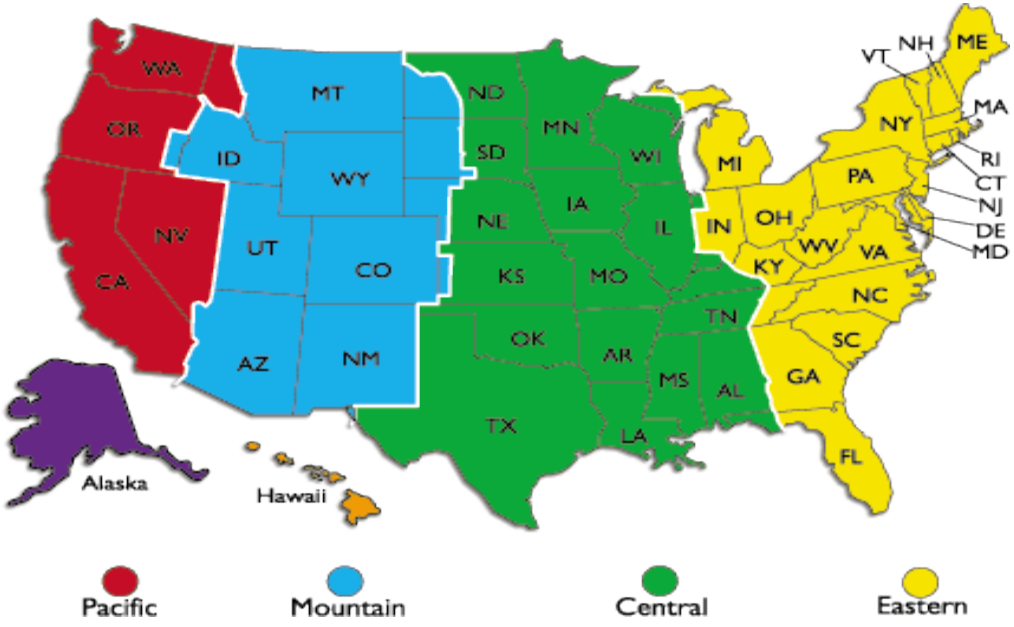
Trends

- http://www.dialidol.com/asp/predictions/DIHardGraphs.asp?week=3&intshow=50&season=8

- Visual confirmation of what was expected based on the percent of the population that is voting at that time

What Does this Mean?

- This means that contestants who are originally from, or popular in the Mountain and the Hawaiian Time Zones, have an advantage



Busy Signals and Saturation

- The toll free phone lines for contestants could be busy at either the regional and the national level
- Suppose that John Smith in Boston calls for her favorite contestant, contestant A, 1000 times but the phone line is at 50% busy signal. Then, only 500 votes get recorded for contestant A.
- Suppose Sally Johnson in Salt Lake City calls for her favorite contestant, contestant B, 1000 times but the phone line is at 5% busy signal. Then, 950 votes get recorded for contestant B.
- American Idol counts votes received…not intended votes

Death by Demographics

- In Season Six, there were three contestants remaining:
- Melinda Doolittle: African-American Motown style soul singer from Nashville (29 years old)
- Jordin Sparks: Half African-American, half Caucasian Pop Singer from Phoenix (17 years old)
- Blake Lewis: Caucasian Urban Rap Singer from Seattle (25 years old)

What Happened

- http://www.dialidol.com/asp/predictions/GeoPredictions.asp?season=6&week=3&sort=Winning
- Dialidol showed an overall busy percentage of 31% for Melinda, 29% for Jordin, and 27% for Blake
- Yet, Melinda was eliminated
- While texting may have hurt her, there is a high probability (given the huge sample size) that Melinda did not have the lowest 'intended vote' but she had the lowest 'tallied vote' and thus went home.

Example

- Consider the following hypothetical scenario:
  - 3 contestants left
- Contestant 1: 28 year old Caucasian male country/blues singer from Richmond, VA
- Contestant 2: 19 year old African American male pop/rock singer from San Diego
- Contestant 3: 24 year old Caucasian female (Mormon) folk singer from Salt Lake City.

Example

Eastern/Central	E/C Attempts	E/C Busy %	E/C Total Counted
Contestant 1	10000000	50%	5000000
Contestant 2	9000000	45%	4950000
Contestant 3	8500000	42%	4930000
Mountain	M Attempts	M Busy %	M Total Counted
Contestant 1	700000	5%	665000
Contestant 2	640000	4%	614400
Contestant 3	820000	6%	770800
Pacific/Alaska	P/A Attempts	P/A Busy %	P/A Total Counted
Contestant 1	1800000	15%	1530000
Contestant 2	2000000	17%	1660000
Contestant 3	1900000	16%	1596000
Hawaii	H Attempts	H Busy %	H Total Counted
Contestant 1	25000	0%	25000
Contestant 2	25000	0%	25000
Contestant 3	25000	0%	25000
Total	ATTEMPTS		TOTAL COUNTED
Contestant 1	12525000		**7220000**
Contestant 2	11665000		7249400
Contestant 3	**11245000**		7321800

Jasmine Trias

- In American Idol Season 3, one of the contestants, Jasmine Trias, was from Hawaii.
- Despite mediocre performances, she finished 3rd that season, ahead of future Grammy and Oscar winner Jennifer Hudson
- The Hawaiian vote was enormous and unlike the rest of the country where less than half of the votes may go through, Hawaii rarely has any busy signals and thus a Hawaiian vote could be worth 2-5 times each mainland vote.

Example

- Consider the following hypothetical scenario:
  - 3 contestants left
- Contestant 1: 27 year old African American female Motown/pop Singer from Jackson, MS
- Contestant 2: 20 year old Caucasian male hard rocker from Beaumont, TX
- Contestant 3: 19 year old Asian/Hawaiian female pop singer from Honolulu, HI

Example

Eastern/Central	E/C Attempts	E/C Busy %	E/C Total Counted
Contestant 1	10000000	50%	5000000
Contestant 2	9500000	48%	4940000
Contestant 3	8500000	43%	4845000
Mountain	M Attempts	M Busy %	M Total Counted
Contestant 1	700000	5%	665000
Contestant 2	710000	5%	674500
Contestant 3	640000	4%	614400
Pacific/Alaska	P/A Attempts	P/A Busy %	P/A Total Counted
Contestant 1	1700000	18%	1394000
Contestant 2	1700000	18%	1394000
Contestant 3	1700000	18%	1394000
Hawaii	H Attempts	H Busy %	H Total Counted
Contestant 1	25000	0%	25000
Contestant 2	25000	0%	25000
Contestant 3	225000	10%	202500
Total	ATTEMPTS		TOTAL COUNTED
Contestant 1	12425000		7084000
Contestant 2	11935000		**7033500**
Contestant 3	**11065000**		7055900

Solutions to the Geography Problem

- Limit to five votes per phone line (and five texts per cell phone) per contestant (like Dancing with the Stars does)
- Weight the final vote by Time Zone
- Adjust the number of votes received per minute as a function of Geography (to try to get similar curves of busy signals in each time zone)
- Allow everyone to vote from 10PM-5AM ET.

- The top solution seems to be the most fair but will never happen as long as AT&T sponsors the show and they make millions of dollars off of these calls.

The Performance Order

- The order that you sing is very important.
- The phone lines do not open until the end of the show…if you sing from 9:52-9:57, you are fresh in the mind of the voters as you “close the show”
- If you sing from 8:05-8:10 and the phone lines don’t open until 10PM, your fans may have forgotten you or gone to sleep or watched something else and by the time voting begins, you’re a distant memory

The Performance Order

- A contestant is assigned their order in the performance by the producers
- While it's not stated in the rules, producers claim that contestants are assigned based on gender (try to alternate boy/girl) or by commonalities in set design for a live show (if two contestants use a piano, they will be put back-to-back to make set changes easier) and that it's not by favoritism.
- Only in the final week is the order determined by a public coin flip.

Adam Lambert

- According to contestants who were voted off, Adam Lambert was given preferential treatment on American Idol season 8 by the producers
- Rumors:
  - He was allowed to get the first choice of song
  - He got to monopolize the time of the band to perfect his arrangement
  - He was allowed to practice after-hours with vocal coaches
  - He got to sing at the end of each show

Adam Lambert

- While we cannot test the other three rumors, we can test the claim that Adam Lambert was given prime positions in the singing order in the 10 weeks leading up to the finals.
- Here is when Adam sang: 11^{th} out of 13, 5^{th} out of 11, 8^{th} out of 10, 8^{th} out of 9, 8^{th} out of 8, 3^{rd} out of 7, 5^{th} out of 7, 5^{th} out of 5, $3^{rd}/4^{th}$ out of 4 (he was in a duet that closed the show), and 3^{rd} out of 3.

Adam Lambert

- He sang last to close the show 3 out of 9 times (4 out of 10 including the duet)
- He sang among the final 1/3 of contestants 8 out of 10 times
- He never sang in the opening 1/3 of all contestants.

Testing a Hypothesis

$$H_0 : p = 0.5$$

$$H_1 : p > 0.5$$

In testing the claim that Adam Lambert sang in the final half of contestants more often than he should have by chance, we have the data that he sang in the final half 8 out of 10 times.

Testing this hypothesis yields a p-value of 0.029. At a 0.05 significance level, we reject the null in favor of the alternative hypothesis.

Indeed, Adam Lambert sang in the final half more often than he should have by chance.

Solutions to this Problem

- Have all contestants each week randomly choose their position in a fair and open way (like the Kentucky Derby pill draw)
- Open the phone lines when the show begins rather than waiting until after the show
  - In conjunction, give every contestant a static phone number that does not change from week to week
  - Dancing with the Stars utilizes these voting techniques.

Solutions to this Problem

- Another Solution is to apply the method of Richard Tapia and his paper: "Some Mathematical Insights Related to BMX Bicycle and Drag Racing “ (2003)
  - That paper examined fair lane assignments in BMX racing and Drag Racing
  - Assign the entire singing order for all weeks before the start of week 1 as an ordered 10-tuplet for each singer
  - Each week delete a new row but everyone else sings as assigned.
  - Works for 10 weeks (week of 12 singers through week of 3 singers)

Sample 10-tuples

Singer	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
A	1	12	6	7	2	11	5	8	3	10
B	2	11	4	9	5	8	3	10	6	7
C	3	10	5	8	4	9	6	7	1	12
D	4	9	1	12	6	7	8	5	10	3
E	5	8	9	4	11	2	12	1	7	6
F	6	7	11	2	3	10	1	12	9	4
G	7	6	12	1	8	5	10	3	11	2
H	8	5	10	3	7	6	2	11	4	9
I	9	4	2	11	1	12	7	6	8	5
J	10	3	8	5	9	4	11	2	12	1
K	11	2	3	10	12	1	4	9	5	8
L	12	1	7	6	10	3	9	4	2	11

Conclusions

- There are numerous ways to make American Idol more fair than it is today
  - But, as long as the ratings are high, producers really could care less about fairness
  - Notice that despite the producer's best efforts, Adam Lambert did not win this season (Kris Allen did).
- Each year the role of text messaging gets greater and greater (and cannot be tracked by Dialidol or any other method) making predicting harder to do.