

> Mining for Major Gifts

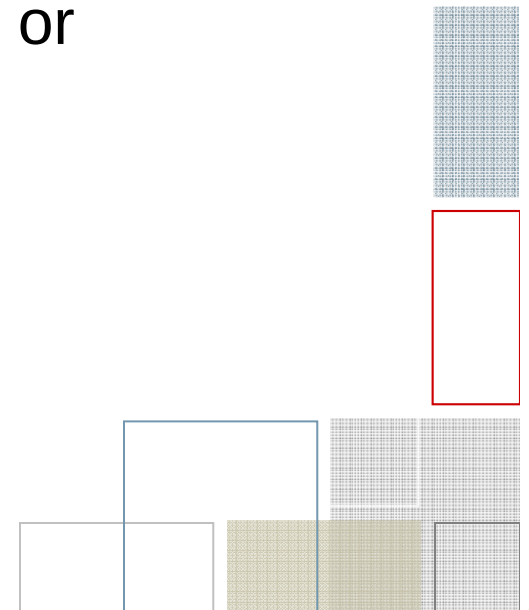
Dan Luperchio
January 28, 2009



> Commonly Asked Questions

SPSS

1. Will I be able to get copies of the slides after the event?
2. Is this web seminar being taped so I or others can view it after the fact?



> Commonly Asked Questions

SPSS

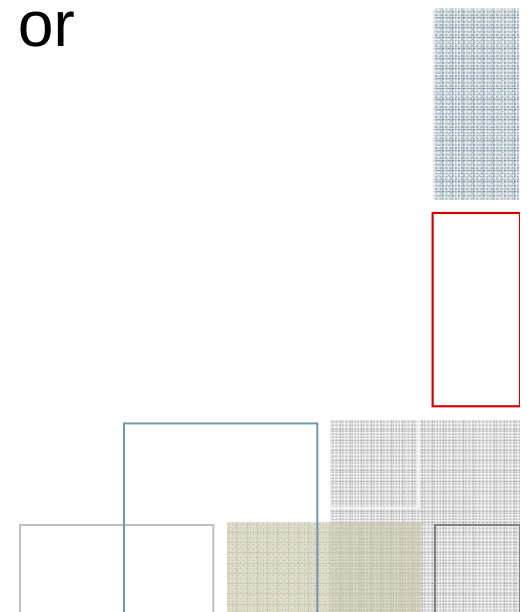
1. Will I be able to get copies of the slides after the event?

Yes

2. Is this web seminar being taped so I or others can view it after the fact?

Yes

www.spss.com/events



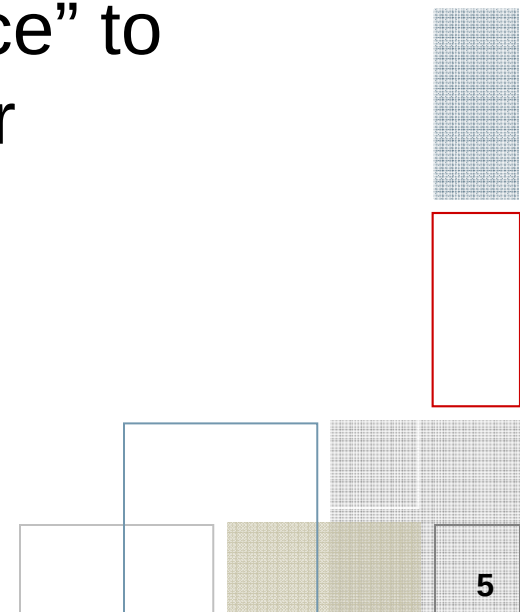
> Table of contents

SPSS

Introduction	slide 2
Data mining concepts	slide 6
Early planning	slide 12
10 Steps to success	slide 16
Predicting major/annual gifts	slide 21
Benefits and implications	slide 23
Acknowledgments	slide 26

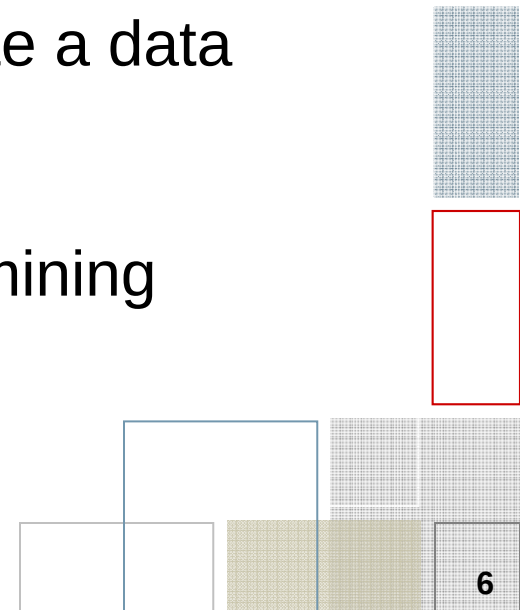
> Primary objectives

- Discuss my experience with data mining for advancement
- Provide concrete steps for starting, executing, and interpreting the results of a data mining project
- Encourage those who are “on the fence” to give data mining a try for their office or program



> One quick question

- Which of the three categories below best characterizes your role in data mining?
 1. **Chief advancement officer.** Decide whether to invest in data mining or commission a data mining project.
 2. **Prospect researcher.** Design and execute a data mining project.
 3. **Fundraiser.** Interpret and act upon data mining results.



> Data mining defined and in practice

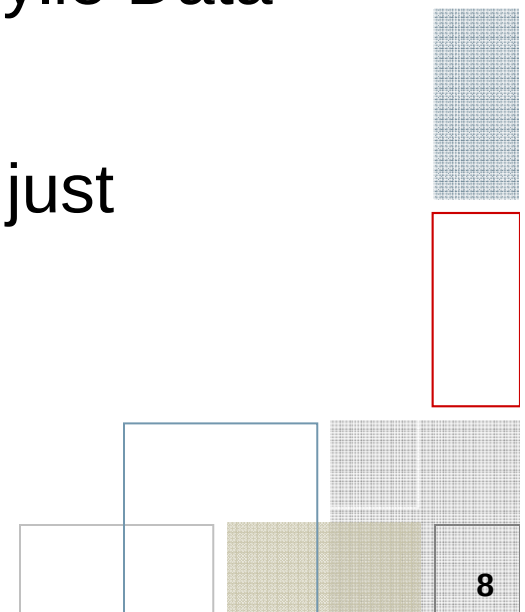
- *“Data processing using sophisticated data search capabilities and statistical algorithms to discover patterns and correlations in large preexisting databases.”*

<http://wordnet.princeton.edu/>

- One major application of data mining is in the business world
 - Stop & Shop, MetLife, Virgin Mobile, BMW and Pep Boys rely on SPSS' Clementine for predictive analytics

> How I arrived in the data mines

- Psychology major at Johns Hopkins University (JHU)
 - Statistics training
- Hired by Arts and Sciences (A&S) development office to lead prospect identification efforts
- Nominated for 2008 Council for Advancement and Support of Education (CASE) Peter B. Wylie Data Mining Internship
- Realized that our prospect research was just *multiple regression* by another name



> What is multiple regression?

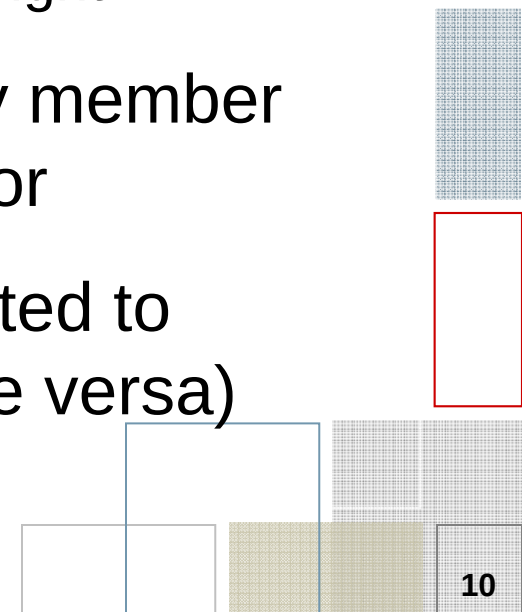
Technically:

“A statistical technique that predicts values of one variable on the basis of two or more other variables.”

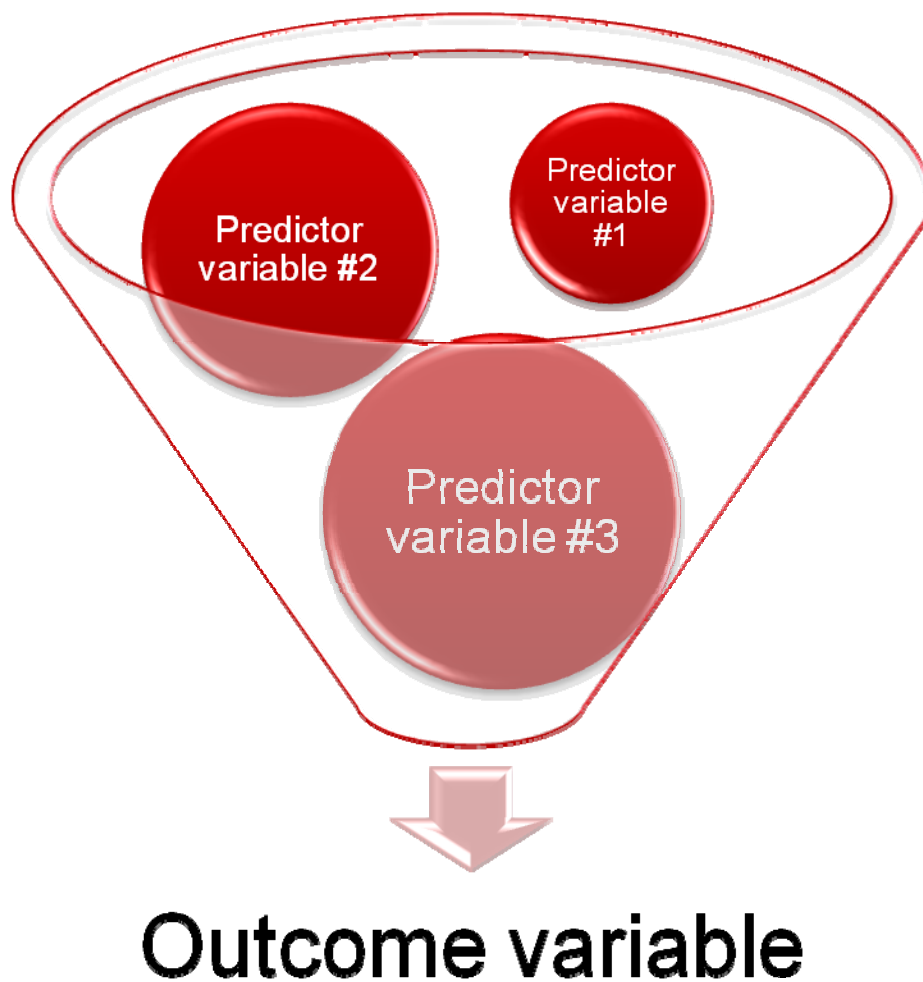
<http://wordnet.princeton.edu/>

> How multiple regression works

- Select an outcome that is likely influenced by several factors
- One by one, assess the unique correlation between these factors and the outcome
- Use these correlations to weight each factor
 - The stronger the correlation, the larger the weight
- Assign these weights as “points” to every member of the population whom exhibits that factor
- Members with the most points are predicted to have the most positive outcome (and vice versa)



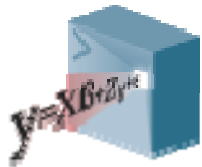
> A visual representation of multiple regression



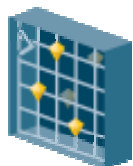
> Finding the software you need



- Affordable technology exists to quickly and easily mine large amount of data
 - Assumes solid preparation, execution, interpretation



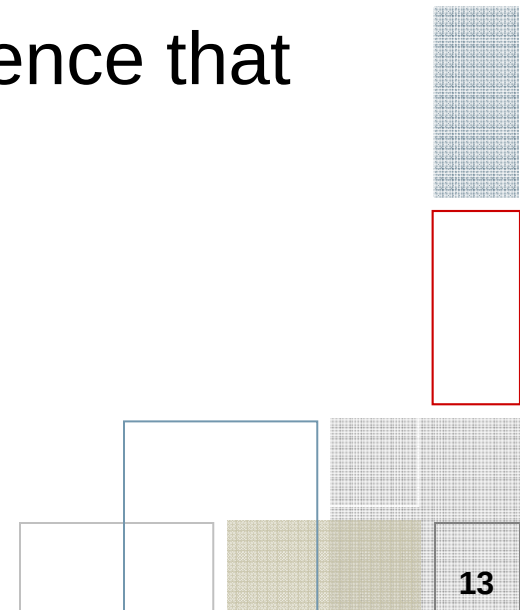
- SPSS Statistics is a suite of products for statistical analysis and data management



- SPSS' Clementine is a data mining workbench that enables institutions to quickly develop predictive models based on large amounts of data from a variety of sources

> Critical questions before getting started

- What do you want to take from this exercise?
- What is your population?
- Given your goals and your population, what is an appropriate outcome variable?
 - Can be either continuous (\$\$) or binary (Y/N)
- What variables could reasonably influence that outcome variable?



> My answers to these critical questions

- What do I want to take from this exercise?
 - A predictive model that identifies new prospects and prioritizes existing prospects for my division
 - Flexible procedure that can be reapplied to other JH divisions
- What is my population?
 - All living alumni of A&S
- What is an appropriate outcome variable?
 - Lifetime giving to JH
- What variables could reasonably influence my outcome variable?

> Variables to consider



- Business address
- City and state of current address
- Country of current address



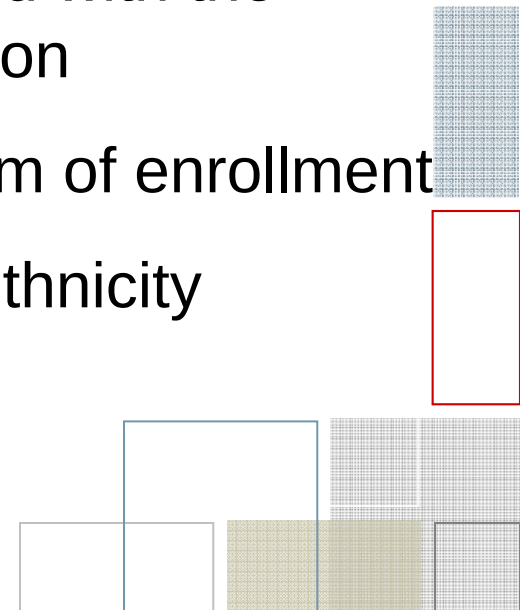
- Graduation year
- Marital status
- Number of children



- Consecutive donor
- Employment specialty
- Gender



- Number of children affiliated with the institution
- Program of enrollment
- Race/ethnicity



> 10 Steps to Success – Preparation

1. Determine your possible predictor variables

- Be creative -- Use common sense, literature, etc.
- Be practical -- Some fields may be inaccessible.



2. Assemble your dataset

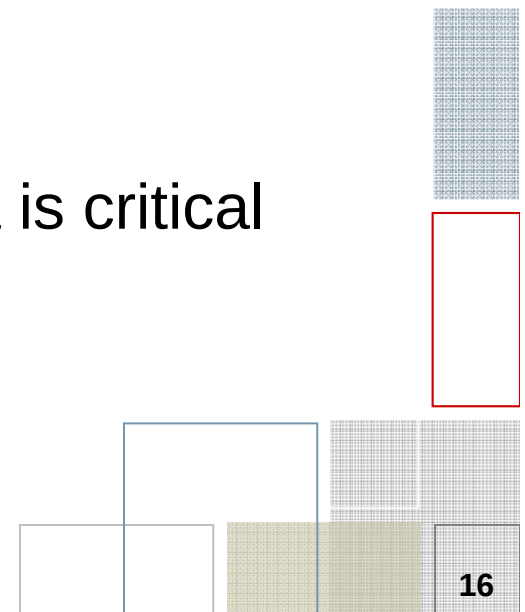
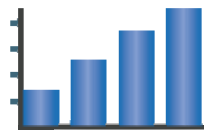
- Data are rarely all wrapped up neatly in one place

3. Clean up your data



4. Review your descriptive statistics

- Understanding the patterns in your data is critical



> 10 Steps to Success – Execution, part 1

5. Test relationships

- Divide your population into samples
- Compare the giving patterns of these samples to the larger population

Lifetime JH Chunks * Gender Crosstabulation

			Gender			Total
			F	M	U	
Lifetime JH Chunks	\$0	Count	8850	12679	41	21570
		% within Gender	57.6%	48.6%	62.1%	52.0%
	\$1-240	Count	4066	5549	14	9629
		% within Gender	26.5%	21.3%	21.2%	23.2%
	\$241+	Count	2456	7840	11	10307
		% within Gender	16.0%	30.1%	16.7%	24.8%
Total		Count	15372	26068	66	41506
		% within Gender	100.0%	100.0%	100.0%	100.0%

> 10 Steps to Success – Execution, part 2

6. Prepare your possible predictors
7. Create your regression model

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	29.161	.862		33.840	.000
	Male Predictor	.000	.012	.000	.022	.983
	Married Predictor	.539	.029	.191	18.420	.000
	Graduation Year	-.015	.000	-.172	-33.877	.000
	Widowed Predictor	.803	.062	.058	12.912	.000
	Business Address Predictor	.401	.012	.140	32.516	.000
	Single Predict	.178	.028	.064	6.331	.000
	Divorced Predict	.393	.047	.041	8.306	.000
	Affiliated Children Predict	.618	.042	.059	14.602	.000
	Known Children Predict	.404	.016	.133	25.929	.000
	Undergrad Predict	.592	.013	.206	44.271	.000
	PT Graduate Predict	.118	.021	.026	5.680	.000
	Northwest Predict	-.105	.039	-.011	-2.719	.007
	Mid Atlantic Predict	.063	.012	.022	5.299	.000
	International Predict	-.245	.030	-.033	-8.211	.000

a. Dependent Variable: Lifetime JH Log

> 10 Steps to Success – Interpretation, part 1

8. Evaluate your model

- Coefficient of determination (R^2)
- Making sense of your model's output
 - Positive, negative, and insignificant predictors
 - All predictors ranked by size of contribution

Statistical model

Outcome variable: Lifetime giving

Strength of model (R^2): 55.1% -- excellent

- For more information about R^2 , often referred to as the coefficient of determination, please consult the *Glossary of Terms*, attached.

Predictor variables used:

Name	Direction	Variable type	Size of Contribution
Graduation year*	Positive	Continuous	Does not compare
Domestic*	Positive	Binary	1.00104
Graduate*	Positive	Binary	0.9476
Canada*	Positive	Binary	0.870968
Married*	Positive	Binary	0.776031
Has business address	Positive	Binary	0.403177

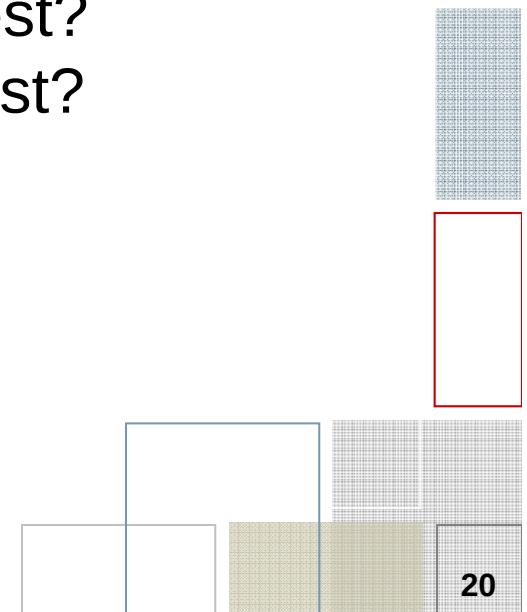
Graduation year squared	Negative	Continuous	Does not compare
Gulf states region	Negative	Binary	-0.170303

*Most influential variables

Predictor variables not used: Mid Atlantic, North Central, Northwest, South Central, Southwest regions

> 10 Steps to Success – Interpretation, part 2

9. Rank your prospective major donors
 - Apply the *Philanthropy Likelihood Index (PLI)* to each individual based on his/her point total
 - Create ten deciles of ascending likelihood to become a major donor
10. Evaluate your model's predictions
 - Are individuals in decile 10 really the best?
 - Are individuals in decile 1 really the worst?



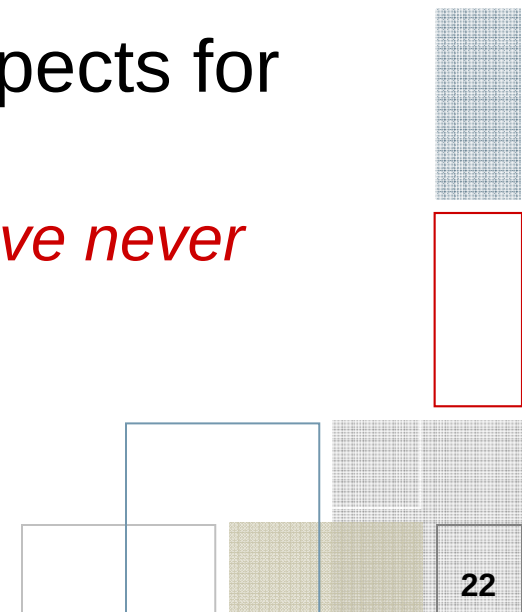
> Predicting major gifts

- I'm not convinced... can the model identify your existing major donors?
 - *Yes! 90% successful for A&S*
- Do development officers with better portfolios (according to this model) raise more money?
 - *Yes! In FY08, FY09 projected, and FY09 actual returns through the halfway mark*
- Did the model identify any new prospects?
 - *Yes! 40% of A&S prospects in decile 10 have never been visited*



> Predicting annual gifts

- Does the model predict annual fund participation?
 - *Yes! In three consecutive mailings, A&S received 10%+ participation from decile 10 and 1-3% from decile 1*
- Does the model predict lead annual gifts?
 - *Yes! In same three mailings, 60-80% of lead gifts were from the models' top donors*
- Does the model identify any new prospects for annual gifts?
 - *Yes! 9% of A&S prospects in decile 10 have never made an annual gift*



> Benefits of the DIY 10-step process

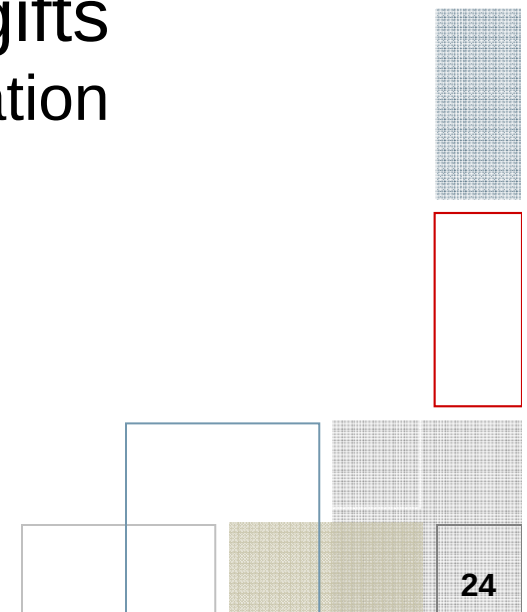
- Universal -- Has already been adopted by ten diverse schools in U.S., Canada, and Europe

Organization type	Country	# Records in data set	R ²	# Variables in formula	% Non-Donors	% Non-Donors in Decile 10	% Major Donors	% Major donors in Decile 10	Median gift in Decile 10	% Total Giving in Decile 10
Private, research university	United States	41,508	36.9%	49	51.9%	8.9%	0.7%	63.2%	\$ 1,200	64.8%
Private, research university	United States	41,508	33.3%	59	66.2%	19.3%	0.5%	72.1%	\$ 520	77.5%
Public, research university	Canada	55,409	36.5%	19	66.8%	12.8%	1.0%	74.0%	\$ 326	80.2%
Private, liberal arts college	United States	24,503	55.1%	10	27.6%	1.1%	1.4%	40.5%	\$ 2,718	44.0%
Private, research university	United States	112,891	38.6%	23	49.7%	0.2%	0.4%	59.8%	\$ 1,250	34.0%
Community college	United States	36,945	46.8%*	10	87.3%	0.5%	1.2%	98.1%	\$ 115	97.6%
Public, research university	England	105,289	7.3%	5	97.2%	83.1%	0.1%	76.2%	\$ -	68.8%
Public university	United States	3,659	18.3%	6	53.2%	25.5%	1.2%	58.5%	\$ 190	68.0%
Public, research university	United States	35,177	18.6%	52	68.2%	35.0%	0.6%	59.6%	\$ 80	55.0%
Land-grant university	United States	190,272	66.5%	28	47.9%	0.0%	0.6%	70.3%	\$ 1,574	76.4%
Private, liberal arts college	United States	19,901	30.3%	24	8.0%	1.7%	2.1%	34.9%	\$ 4,188	37.8%
Average	-	60,642	34.1%	26	56.7%	17.1%	0.9%	64.3%	\$ 1,106	64.0%

- Testable, repeatable
- **Inexpensive!**

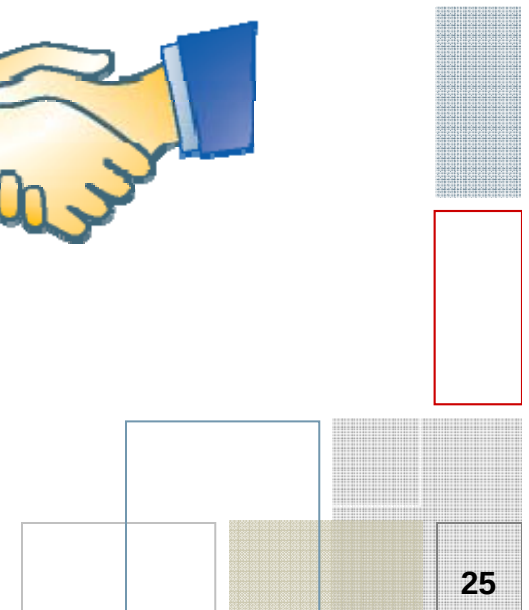
> Benefits of the predictive model

- All individuals rated 1-10 on basis of likelihood to become major/annual donor
- All predictor variables positively, negatively, and not linked with the outcome variable
 - Ranked according to size of contribution to model
- Specific predictions for major/annual gifts
 - Ranking of existing prospects for prioritization
 - List of new prospects for major gifts
 - List of new prospects for annual gifts



> Ideas for future projects

- Specific constituencies
 - Part-time graduate alumni
 - Undergraduate parents
 - Expanding the set of possible predictors
- Specific fundraising initiatives
 - Athletics
- Alternative giving vehicles
 - Planned giving



> Acknowledgments

- Peter B. Wylie
 - Peter is a national expert and leading author in the field of data mining.
- CASE
 - Chris Thompson and the CASE Research staff advised on the data mining process and creation of the model.
- SPSS
 - SPSS Inc. is a leading global provider of predictive analytics software and solutions.

> Any questions?

SPSS

- Download executive summary or white paper
- Attend CASE District II conference
 - March 22-24, 2009
 - Baltimore, MD
 - Visit <http://casetwo.org/> for details and registration info
- Email dal@jhu.edu