Analyses in Reproducible Research:

An Example with Logistic Regression

Ryan T. Pohlig

Presentation Overview

- Logistic Regression
 - Overview
- Brief Intro to SPSS
 - Using Syntax in SPSS
- Example Research
 - Quick Overview
 - Talk about the data
- Running the analysis in SPSS
 - Checking assumptions
 - Interpreting results

- DV and Outcome are going to be used interchangeably
- IV and Predictor will be used interchangeably

Logistic Regression

Crash Course

Logistic regression

- Conceptually it is no different than the multiple egression you (may) be familiar with
- Research Questions are even the same
 - Does an IV predict a DV?
 - Does a set of IVs predict a DV?
 - Does an IV predict a DV, after adjusting for covariates?
- Differences:
 - When an DV is categorical, cannot use multiple regression
 - Interpretation of coefficients are different

Categorical DVs

- Different Methods for analysis given a categorical DV
 - Discriminant Function Analysis
 - Part of MANOVA family (IV's need to be interval or ratio)
 - Loglinear Modeling (Multi-way Frequency Analysis)
 - Generalization of two-way chi-square test (or more dimensions than Mantel-Haenszel test)
 - Generalized Linear Modeling
 - Generalization of the more familiar General Linear Modeling (ANOVA/Regression)
 - Specify a link function & a distribution from the exponential family
 - More flexible, can accommodate many types of outcomes
 - Probit Regression or Logistic Regression
 - Poisson Regression

LR : Regression

- Predictors & IVs function the same as in regression
 - Can have continuous predictors
 - Categorical ones must be dummy or contrast coded
 - Can be sequential with tests for multiple blocks of predictors
 - Can have Automated selection procedures
 - Can test moderation using interactions

- Technically LR is regression looking at a non-linear relationship between predictors and outcomes
- One can linearize arelationship by
 - Applying a transformation to the outcome
 - Applying a transformation to the predictors

LR : MR part 2

- MR formula in matric notation
 - $y = \beta X + \varepsilon$
 - *y* n by 1 vector of observed DV
 - β p by 1 vector of regression coefficients
 - X n by p matrix of observed IVs
 - ε n by 1 vector of error terms
- Applying transformation to the outcome
 - $g(y) = \beta X + \varepsilon$
- Applying Transformation to predictors & errors

• $y = f(\beta X + \varepsilon)$

• In LR, the logit function is used

•
$$g(y) = logit = log\left(\frac{\pi}{1-\pi}\right)$$

- Where π is the probability y=1 given your predictors
 - Formally: $\pi = P(y = 1 | X = x)$
- Where 1-π is the probability y=0 given your predictors

• Formally: $1 - \pi = P(y = 0 | X = x)$

LR Model

• In logit form

•
$$g(y) = \log\left(\frac{\pi}{1-\pi}\right) = X\beta + \varepsilon$$

• Solve for $\boldsymbol{\pi}$

•
$$\pi = \frac{e^{x\beta + \varepsilon}}{1 + e^{x\beta + \varepsilon}}$$

• For predicted model

•
$$g(\hat{y}) = \log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = X\hat{\beta}$$

 \bullet Solve for π

•
$$\hat{\pi} = \frac{e^{x\hat{\beta}}}{1+e^{x\hat{\beta}}}$$

- In order to solve for $\hat{\beta}$, need to use maximum likelihood
 - Iterative process that maximizes the sample values to represent the population parameters
- Likelihood probability density function

•
$$\ell(\hat{\beta}) = \prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1 - \hat{y}_i)}$$

• Log-likelihood function

•
$$L(\hat{\beta}) = \log(\ell(\hat{\beta})) =$$

 $\sum_{i=1}^{n} [y_i \log(\hat{\pi}_i) + (1 - y_i) \log(1 - \hat{\pi}_i)]$

LR

- Using this you can get a global likelihood ratio statistic for testing the entire model
 - $L(\hat{\beta})$ can be used in the log-likelihood ratio test
- Also individual predictors can be tested using a Wald chi-square test
 - $W_j = \left[\hat{\beta}_j / SE(\hat{\beta}_j)\right]^2$, where j refers to the predictor of interest
- LR Assumptions:
 - Independence
 - No multicollinearity
 - No outliers/influential cases
 - Linearity of continuous IV's with logit

Example

	Smoke											
		Νο	Yes	Total								
Weight	Norm	223	114	337								
	Low	70	81	151								
	Total	293	195	488								

- DV = Birth Weight
 - Low or Not
- IV = Mother Smoking during pregnancy
 - Yes or No

• Odds

• Prob. of Low Weight for non-smokers?

•
$$\frac{\frac{70}{293}}{\frac{223}{293}} = \frac{.239}{.761} = .314$$

- 31.4% chance of low birth weight for nonsmokers
- For smokers?

•
$$\frac{\frac{81}{_{195}}}{\frac{114}{_{195}}} = \frac{.415}{.585} = .711$$

- 71.1% chance of low birth weight for smokers
- Odds ratio
 - $\widehat{OR} = \frac{.314}{.711} = 2.26$
 - 2.3 times more likely to have a low birth weight, if the mother smoked during pregnancy

Logistic Regression Results

- Logistic Regression results
 - Given you coded smoking as 0 for no, 1 for yes
 - $\widehat{Logit} = -1.159 + .817(smoke)$
- The betas are not interpreted directly
- β_j is amount of change in logit for each one unit change in X_j
- The null-hypothesis is essentially testing if the Odds Ratio differs than 1

- In order to make the parameter estimates interpretable, we need to exponentiate them
 - Intercept is probability of low birth weight given no smoking

•
$$e^{\hat{\beta}_0} = e^{-1.159} = .314$$

• Full equation is probability of low birth weight given smoking

•
$$e^{\widehat{\beta}_0} = e^{-1.159 + .817} = .710$$

 Slope is Odds Ratio, the impact of smoking on birth weight

•
$$e^{\hat{\beta}_1} = e^{.817} = 2.264$$

LR- is your model good?

- There are different estimates for in R^2 LR (yes more than one way exists)
- Psudeo R^2 are used
 - http://stats.idre.ucla.edu/other/multpkg/faq/general/faq-what-are-pseudo-rsquareds/
- Cox & Snell
 - Looks at improvement of predictors relative to a null model (intercept only)
- Negelkerke
 - Adjustment to Cox & Snell so that the max value can be 1
- Can look at Information Criteria to determine fit
 - AIC, BIC, etc.
 - Smaller of these indicates best fit

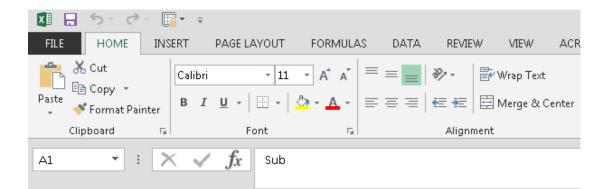
- Can calculate predicted probabilities for individuals' or use regression formula for a specific set of X values
- Hosmer-Lemeshow is another method for showing goodness of fit
- Many Use an ROC curve for significant predictors
 - Can look at Area Under the Curve
 - Youden's Index or Kullback-Leibler distance
 - Maximum value of Youden's index is the cut point for classification that minimizes both false negatives & false positives
 - Is Sensitivity + (Specificity-1)

SPSS

Brief Intro

Check Data file

- Can Open .csv files with excel pretty easily
- Notice we have a header row with variable names
- Open SPSS
 - Choose File -> Open -> Data
 - Navigate to the correct folder
 - Make sure to choose .csv as file type
 - Hit 'open'



	А	В	С	D	E	F	G	Н	I	
1	Sub	ag	at	as	an	ai	eag	eat	eas	е
2	2	2.22	0.354	0.58	0.686	0.601	1.267	0.336	0.346	
3	43	2.681	0.475	0.672	0.868	0.667	2.053	0.44	0.52	
4	25	1.979	0.343	0.508	0.624	0.504	1.2	0.299	0.396	
5	65	1.747	0.269	0.476	0.525	0.476	0.612	0.147	0.017	
6	70	2.99	0.599	0.686	1.039	0.667	2.513	0.543	0.607	
7	16	2.917	0.483	0.763	0.901	0.77	2.2	0.462	0.637	

ta Open Data								X
Look <u>i</u> n: 🚺	Reproducible Resear	rch	▼ 😰) 🔯 🎟 -				
SPSS	ncer.csv a.csv							
File <u>n</u> ame:	glaucoma.csv							
	giaucoma.csv							<u>O</u> pen
Files of type:	CSV (*.csv)						-	<u>P</u> aste
<u>E</u> ncoding:	Unicode (UTF-8)						-	Cancel <u>H</u> elp
			Ret	rieve File From	Repository		_	

Importing Data

- Point & Click
 - SPSS does a good job at naively having the correct settings
 - Choose Delimited
 - With a CSV file, a comma is the delimiter
 - Make sure the top row is the variable names
 - Hit Next, Hit Next for Step 3
 - Make sure to choose Comma in step 4
 - Hit next all the way through to go right to the data
 - Alternatively choose to 'paste syntax'
 - This opens a syntax file that can be used to skip the import data wizard in the future (Please save the syntax file now)

Text Import Wizard - Step 2 of 6				×			
How are your variables arranged? Delimited - Variables are delir DEixed width - Variables are aligr			i.e., comma, ta	ab).			
Are variable names included at the top Yes Line number that contains variabl No		4					
What is the decimal symbol? Period Comma							
Text file: C:\Users\Ryan Pohlig\Deskto 0	20	30					
Name Sub, ag, at, as, an, at 2 2,2.22,0.354,0.58 4	Which del	nt Wizard - Deli imiters appear a na nd Trailing Spa emove leading emove trailing s	imited Step 4 between vari Semicolon aces spaces from	of 6 ables? string values	◉ N <u>o</u> ◯ Sir	igle <u>q</u> uote uble quote	lifier?
	Sub 2 43 25 65 70 16 6 5 12 63 1	ag 2.22 2.681 1.979 1.747 2.99 2.917 2.228 1.835 2.174 3.153	at 0.354 0.475 0.343 0.269 0.599 0.483 0.355 0.312 0.365 0.53	as 0.58 0.672 0.508 0.476 0.686 0.763 0.601 0.463 0.57 0.834	an 0.686 0.868 0.624 0.525 1.039 0.901 0.686 0.578 0.668 0.976	ai 0.601 0.667 0.504 0.476 0.667 0.77 0.586 0.483 0.572 0.814	eac 1.2 2.0 1.2 0.6 2.5 2.2 1.0 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0
		< <u>B</u>	ack <u>N</u> ext ≍	Finish	ancel He	alb	

SPSS has 3 main Windows

- Data Window has 2 tabs
- The Data View
 - Looks like an excel file of your data
 - Contains your actual raw data
 - Notice the button at the top that has arrows pointing to an 'A' and a '1'
 - This button is used for showing variable labels
 - Do you want to see what the raw numeric values (0 or 1)
 - Or what those raw values stand for (Male or Female)

		5 2			H H					ARS								
			0.		the set												Visible:	64 of 64 Va
	🖉 Sub	🥒 ag	🥒 at	🥒 as	🥒 an	ar 🕼	🦧 eag	17	🧳 eas	🧳 ean	🥒 eai	🖉 abrg	🥒 abrt	🥖 abro	🧳 abrn	🧳 abn	/ hic	🖉 mhci
1	2	2.220	364	580	.686	601	1.267	356	.346	256	.331	,479	.260	107	.014	.098	.214	
2	43	2.681	,475	.672	.868	.667	2.053	.440	.520	.639	,454	1.090	.377	.257	.212	.245	.382	.1
3	25	1.979	343	.608	.624	.604	1.20	.299	.396	.259	.246	466	.209	.112	.041	103	.195	
4	65	1.747	.269	.476	.525	.476	.612	.147	.017	.044	.405	.170	.062	.000	.000	108	030	0
5	70	2.990	599	.686	1.039	.687	2.613	.543	.607	.871	.492	1.800	.431	.494	601	.274	.383	1
6	16	2.917	.483	.763	.901	,770	2.200	.462	.637	.504	.597	1.311	.394	.365	.251	.301	.442	.1
7	6	2.228	.365	.601	.686	.586	1.081	298	.243	.153	.387	482	.188	.102	033	159	.180	.0
8	5	1.835	.312	.463	.578	.483	.687	.226	.247	.023	.192	.214	.126	.040	.000	.040	023	0
9	12	2.174	.365	.578	668	.672	.865	.253	.353	038	.221	.140	.073	.033	.000	.034	- 005	- 0
10	63	3.153	.530	.834	.976	.814	2.099	.472	.534	.517	.576	.709	.162	.163	.197	.186	.181	.0
11	11	1.987	308	534	.696	549	1.007	.272	. 369	.034	.332	.345	.190	.065	.000	.090	.076	0
12	62	2.137	.382	.523	.688	.544	1.302	.326	.319	.229	.427	.343	.132	.090	.012	.109	.071	.0
13	81	2.591	.408	.686	792	705	2.010	383	496	529	601	1 253	339	330	210	.375	394	1
14	97	2.198	.306	.549	.711	.551	1.247	.224	.366	.395	.263	.573	.124	.140	.170	.139	.110	.0
15	15	2.620	.450	.658	836	676	1,718	434	426	304	.554	784	.312	192	.095	185	.206	.0
16	4	2.309	.307	.607	.715	.600	1.276	.368	.491	,196	.222	.669	.206	.255	.036	.093	.169	.0
17	14	1.686	.243	463	.488	.492	.420	150	.043	.026	200	.084	.004	.000	.000	.000	- 189	1
18	23	1.490	.262	.305	.468	.376	.415	.166	.103	.021	.125	.054	.037	.010	.000	.007	029	+.0
19	46	2.609	.435	.673	.816	.685	1.935	.391	.461	.533	.650	1.212	.320	.311	.270	.310	283	.1
20	56	2.071	.334	.530	.633	.574	1.366	.317	.327	.317	.406	.636	.241	.159	.044	.191	.297	.1
21	71	2.890	.487	.754	.916	.734	2.275	.448	.680	.694	.552	1.691	.374	.437	.535	.346	.366	.0
22	83	3.536	.635	.077	1.145	.879	3.058	.619	.773	.920	.746	1.994	.561	.449	.515	.469	.311	.0
23	10	2.465	,425	.636	.795	.608	2.019	.384	.463	.625	.547	1.020	.322	.258	.176	.264	.336	.1
24	26	2.201	.413	.581	.726	.561	1.471	296	.353	.399	.423	.737	.181	.205	.181	.170	.263	.1
25	9	1.893	.327	.488	.595	.483	1.133	.273	.319	.234	.306	.408	.146	.078	.025	.159	.185	.0
26	45	2.070	.383	.500	.695	.492	.746	.234	.138	.197	.177	.242	.146	.030	.014	.052	.096	.0
27	78	3.482	.644	.842	1.137	.859	2.315	.579	.719	.604	.413	2.455	.550	.785	.576	.544	.185	.0
28	90	2.999	.554	.724	.905	.737	2.405	.533	.626	.697	.549	1,422	.393	.323	.372	.335	.324	.0
29	94	4.106	.705	1.028	1.319	1.055	3.705	.660	.962	1.218	.865	2.887	.540	.724	1.093	.530	.329	.0
30	13	1.852	.296	.509	.560	.487	.512	.152	.135	.038	.187	.198	.097	.061	.025	.014	.005	.0
31	84	2.140	.387	.547	.684	.523	1.520	.352	.462	.308	.399	.614	.269	.142	.086	.118	.257	.0
32	57	2.068	.357	.539	.653	.519	.871	.254	.231	.120	.266	.141	.108	.020	.000	.013	.061	.0
33	60	2.808	.494	.691	.890	.732	1.452	.436	.456	.145	.415	.618	.322	.136	.011	.149	.052	0
34	28	1.508	.235	.406	.460	.408	.695	.168	.138	.050	.239	.038	.027	.000	.000	.011	043	0
-	101	2 207	700	007	1.0.47	000	0.410		C70	400	7/0	1.000	470	242	247	170	222	

SPSS' Windows

- Data Window has 2 tabs
- The Variable View
 - Contains information for each variable
 - Here you can label Variables
 - Label values (name what the indicators mean)
 - Alternatively You can do this in syntax
- You will notice a variable at the end called 'Class' that is not numeric

	🖨 🗉] 🗠 🤈	¥ 🎬			× 5				45	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Sub	Numeric	4	0		None	None	8	🗮 Right	🛷 Scale	🖒 Input
2	ag	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
3	at	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
4	as	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
5	an	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
6	ai	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
7	eag	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
8	eat	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
9	eas	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
10	ean	Numeric	5	3		None	None	8	🚎 Right	🛷 Scale	🦒 Input
11	eai	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
12	abrg	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
13	abrt	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
14	abrs	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🖒 Input
15	abrn	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
16	abri	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
17	hic	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
18	mhcg	Numeric	6	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
19	mhct	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
20	mhcs	Numeric	6	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
21	mhcn	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
22	mhci	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
23	phcg	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
24	phct	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
25	phcs	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
26	phon	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
27	phci	Numeric	6	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
28	hvc	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
29	vbsg	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
30	vbst	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
31	vbss	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🦒 Input
32	vbsn	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🔪 Input
33	vbsi	Numeric	5	3		None	None	8	🗮 Right	🔗 Scale	🦒 Input
34	vasg	Numeric	5	3		None	None	8	🗮 Right	🛷 Scale	🔪 Input
35	vast	Numeric	5	3		None	None	8	🚟 Right	🖋 Scale	🔪 Input
36	vass	Numeric	5	3		None	None	8	🗃 Right	🔗 Scale	🔪 Input
_	1		_								

SPSS' Windows

- Output windows displays the results/output and contains any error messages
- Syntax window is the part you can directly manipulate code
- Comments will appear in gray on the left while commands appear in black

In Ent Yiew Quada Transform Graphic Ubited Mathematic Transform Graphic Heip Image: Strip Economic <			• X
Image: Construction	BM SPSS Statistics Syntax Edito		
Image: Second	<u>File E</u> dit <u>V</u> iew <u>D</u> ata <u>T</u> r	ransform <u>A</u> nalyze Direct <u>M</u> arketing <u>G</u> raphs <u>U</u> tilities <u>R</u> un Tool <u>s</u> Extensions <u>W</u> indow <u>H</u> elp	
Image: Second			
PRESENCE 1 PRESERVE. 1 ET 3 SET DECIMAL DOT. ************************************			
PRESENCE 1 PRESERVE. 1 ET 3 SET DECIMAL DOT. ************************************	🖌 🏑 🖊 😽	🟹 🎸 🛑 🍼 📖 🖳 🚺 Active: DataSet1 🔽	
IFI TAM 2 PRESERVE. IED TAM SET DECMAL DOT. RESTOR. 4 XORE. 6 GET DATA //YPE=TXT SECURE. 6 FLETC'.VARRPR pro Polig/Desktop/Lectures/Presentations/Reproducible Research/glaucoma.cov* WARSET IMME 7 FLECOV.VARRPS-TAM Polig/Desktop/Lectures/Presentations/Reproducible Research/glaucoma.cov* WARSET IMME 6 0 FLDATERS=''. 10 ARRANGEMENT-DELIMITED 11 0 FLDATERS=''. 12 0 FLDATERS=''. 13 MARNOEMENT-DELIMITED 14 Sub AUTO 15 9 AUTO 16 9 AUTO 17 SA AUTO 18 9 AUTO 19 9 AUTO 19 9 AUTO 19 9 AUTO 10 9 AUTO 114 Sub AUTO 115 9 AUTO 116 9 AUTO 117 S AUTO 118 9 AUTO 119 9 AUTO 120 S AUTO 121 Stat AUTO	PRESERVE.		
HET DATA 3 ST DECIMAL DOT. RESTORE, ADLE, 6 TOTATA //YPE=TXT RECUTE 6 COLONIG-ENTRP NATESET NAME 7 ECTORE, MERSET NAME 7 ECTORE, MERSET NAME 7 NATESET NAME 7 CELONE 7 CELONE 7 NATESET NAME 7 CELONE 7 CELONE 7 NATESET NAME 7 CELONE 7 7 7 7 7 NATESET NAME 7 CELONE 7 <th7< th=""> <th7< th=""> 7</th7<></th7<>	SET	2 PRESERVE.	
CACHE. S C CET DATA //YPE=NT 8000000000000000000000000000000000000	GET DATA		
SEE-UTE 6 #/TLE=*C.WsensRyan PohingDesktopLecturesPresentationsReproducible Researchiglaucoma.csv* WATASET NAME 7 //RX00000-VTTF87 0 DELCASE=LINE 0 10 //RRANGEMENT=DELMITED ///RX100 11 //RSTCASE=2 ///RSTCASE=2 12 //DATATYEENN PERCENTAGE=95.0 ////RX10 13 //ARRAELES= ///RX10 14 Sub AUTO ///RX10 15 ag AUTO ///RX10 16 at AUTO ///RX10 17 as AUTO ///RX10 18 an AUTO ///RX10 19 al AUTO ///RX10 19 al AUTO ///RX10 19 al AUTO ///RX10 10 eag AUTO ///RX10 11 eag AUTO ///RX10 12 eag AUTO ///RX10 13 an AUTO ///RX10 14 eag AUTO ///RX10 15 abr AUTO ///RX10 16	RESTORE.		
WATASET NAME 7 // ENCODING-UTF6' 0 OPELCASE-LINE 0 10 // ARRANCEMENT=DELIMITED // ARRANCEMENT=DELIMITED 11 // RISCASE=2 // ARAILES= 12 // ARAILES= // ARAILES= 14 Sub AUTO // ARAILES= 15 a AUTO // ARAILES= 16 at AUTO // as AUTO 18 an AUTO // as AUTO 19 ai AUTO // as AUTO 21 eas AUTO // as AUTO 22 eas AUTO // as AUTO 23 ean AUTO // as AUTO 24 eai AUTO // as AUTO 25 abr AUTO // as AUTO 26 abr AUTO // as AUTO 27 abr AUTO // as AUTO 28 abr AUTO // as AUTO 29 abr AUTO // as AUTO 29 abr AUTO // ar AUTO 30 hic AUTO // ar AUTO 31 mic AUTO<	CACHE.		
8 OPELCASS=LUNE 9 ODELMTED 11 /ARRANGEMENT=DELINITED 12 /ARRANGEMENT=DELINITED 13 /ARRANGEMENT=DELINITED 14 /ARRANGEMENT=DELINITED 15 /ARRANGESS 14 /ARRANGESS 15 /ARRANGESS 16 at AUTO 17 as AUTO 18 as AUTO 19 ai AUTO 19 ai AUTO 10 as AUTO 11 as AUTO 12 as AUTO 13 as AUTO 14 as AUTO 15 as AUTO 16 as AUTO 17 as AUTO 18 ath AUTO 29 as AUTO 21 as AUTO 22 as AUTO 23 ash AUTO 24 ash AUTO 25 ash AUTO 26 ash AUTO 27 ash	EXECUTE.	6 /FILE="C:\Users\Ryan Pohlig\Desktop\Lectures\Presentations\Reproducible Research\glaucoma.csv"	
9 OPELIMITERS	DATASET NAME		
10 /ARRANGEMENT=DELMITED 11 //ARRAUGEMENT=DELMITED 12 /DATATYPEMIN PERCENTAGE=95.0 13 //ARRAULES= 14 Sub AUTO 15 ag AUTO 16 at AUTO 17 as AUTO 18 an AUTO 19 at AUTO 19 at AUTO 20 eag AUTO 21 eas AUTO 22 eag AUTO 23 eag AUTO 24 eag AUTO 25 eag AUTO 26 eag AUTO 27 abar AUTO 28 abar AUTO 29 abar AUTO 20 eag AUTO 21 eag AUTO 22 eag AUTO 23 abar AUTO 24 eai AUTO 25 abar AUTO 26 abar AUTO 27 abar AUTO 28 abar AUTO 30 mbcg AUTO 31 mbcg AUTO 32			
11 FIRSTCASE=2 12 /DATATYPEMIN PERCENTAGE=95.0 13 NARABLES= 14 Sub AUTO 15 ag AUTO 16 at AUTO 17 as AUTO 18 an AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abtr AUTO 26 abtr AUTO 27 abtr AUTO 28 abtr AUTO 29 abtr AUTO 21 eas AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abtr AUTO 26 abtr AUTO 27 abtr AUTO 28 abtr AUTO 39 abtr AUTO 31 mheg AUTO 32 mhes AUTO 33 mhes AUTO 34 mhen AUTO 35 mhes AUTO <th></th> <th></th> <th></th>			
12 /DATATYPENIN PERCENTAGE=95.0 13 /VARABLES= 14 Sub AUTO 15 ag AUTO 16 at AUTO 17 as AUTO 18 an AUTO 19 ai AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 eas AUTO 26 eai AUTO 27 abrt AUTO 28 abrt AUTO 29 abrt AUTO 29 abrt AUTO 20 eas AUTO 21 ear AUTO 22 eas AUTO 38 abrt AUTO 39 abrt AUTO 31 mbcg AUTO 32 mbc AUTO 33 mbc AUTO 34 mbc AUTO 35 mbc AUTO 36 mbc AUTO 37 pbc AUTO 38 mbc AUTO <th></th> <th></th> <th></th>			
13 //APIABLES= 14 Sub AUTO 15 ag AUTO 16 at AUTO 17 as AUTO 18 an AUTO 19 ai AUTO 20 esq AUTO 21 est AUTO 22 ess AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abr AUTO 28 abr AUTO 29 abri AUTO 31 mbcg AUTO 32 mbc AUTO 33 mbc AUTO 34 mbc AUTO 35 mbc AUTO 36 pbc AUTO 37 pbc AUTO 38 pbc AUTO 39 mbc AUTO 31 mbc AUTO 32 mbc AUTO 33 mbc AUTO 34 mbc AUTO 35 <th></th> <th></th> <th></th>			
14 Sub AUTO 15 ag AUTO 16 at AUTO 17 as AUTO 18 an AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eal AUTO 25 abra AUTO 26 abra AUTO 27 abra AUTO 28 abra AUTO 29 abra AUTO 20 abra AUTO 21 eat AUTO 22 eas AUTO 23 abra AUTO 24 eai AUTO 25 abra AUTO 26 abra AUTO 27 abra AUTO 28 abra AUTO 30 hic AUTO 31 mhcg AUTO 32 mhch AUTO 33 mhcs AUTO 34 mhch AUTO 35 mhci AUTO 36 phcs AUTO 37 phct AUTO			
16 a AUTO 17 as AUTO 18 an AUTO 18 an AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abrg AUTO 27 abrs AUTO 28 abrg AUTO 29 abrd AUTO 20 abrg AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abrg AUTO 27 abrs AUTO 28 abr AUTO 30 hic AUTO 31 mkcg AUTO 32 mkch AUTO 33 mkcs AUTO 34 mkci AUTO 35 mkci AUTO 36 pkcg AUTO 38 mkci AUTO 38 pkcs AUTO			
16 ar AUTO 17 as AUTO 18 an AUTO 19 ai AUTO 20 eag AUTO 21 eaf AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abr AUTO 26 abr AUTO 27 abrs AUTO 28 abr AUTO 29 abr AUTO 29 abr AUTO 29 abr AUTO 30 hic AUTO 31 mhct AUTO 33 mhct AUTO 34 mhci AUTO 35 mhci AUTO 36 phc AUTO 37 phc AUTO 38 mhci AUTO 39 binc AUTO 30 hic AUTO 33 mhci AUTO 34 mhci AUTO 35 mhci AUTO 36 phci AUTO 37 phci AUTO 38 phci AUTO 39 phci AUTO <tr td=""></tr>			
17 as AUTO 18 an AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abtr AUTO 26 abtr AUTO 27 abtr AUTO 28 abtr AUTO 29 abtr AUTO 30 hic AUTO 31 mhcg AUTO 32 mhct AUTO 33 mhct AUTO 34 mhct AUTO 35 mhci AUTO 36 phct AUTO 37 phct AUTO 38 phcs AUTO			
18 an AUTO 19 ai AUTO 20 eag AUTO 21 eat AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abrd AUTO 27 abrs AUTO 28 abr AUTO 29 abr AUTO 28 abr AUTO 29 abr AUTO 29 abr AUTO 30 hic AUTO 31 mhcg AUTO 32 mhci AUTO 33 mhcg AUTO 34 mhci AUTO 35 mhci AUTO 36 phcg AUTO 37 phcs AUTO 38 phcs AUTO 38 phcs AUTO 38 phcs AUTO			
19 ai AUTO 20 eag AUTO 21 eat AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abtr AUTO 26 abt AUTO 27 abrs AUTO 28 abtr AUTO 29 abr AUTO 31 mhcg AUTO 32 mhch AUTO 33 mhcs AUTO 34 mhch AUTO 35 mhck AUTO 36 abtTO 37 phcs AUTO 38 phcs AUTO 38 phcs AUTO			
21 ear AUTO 22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abr AUTO 27 abrs AUTO 28 abm AUTO 29 abr AUTO 28 abm AUTO 29 abr AUTO 30 hic AUTO 31 mhcg AUTO 32 mhct AUTO 33 mhcs AUTO 34 mhcn AUTO 35 mhci AUTO 36 phcs AUTO 37 phcs AUTO 38 phcs AUTO			
22 eas AUTO 23 ean AUTO 24 eai AUTO 25 abrg AUTO 26 abtr AUTO 27 abrs AUTO 28 abm AUTO 29 abr AUTO 30 hic AUTO 31 mhcg AUTO 32 mhc AUTO 33 mhcs AUTO 34 mhcn AUTO 35 mhci AUTO 36 phcg AUTO 37 phcs AUTO 38 phcs AUTO		20 eag AUTO	
23ean AUTO24eai AUTO25abrg AUTO26abrt AUTO27abrs AUTO28abm AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phcs AUTO38phcs AUTO38phcs AUTO39hic AUTO		21 eat AUTO	
24eai AUTO25abrg AUTO26abrt AUTO27abrs AUTO28abr AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcs AUTO37phct AUTO38phcs AUTO39phcs AUTO			
25abrg AUTO26abrt AUTO27abrs AUTO28abm AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhci AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39phcs AUTO			
26abr AUTO27abrs AUTO28abm AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39phcs AUTO			
27abrs AUTO28abm AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhcs AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39mot AUTO			
28abm AUTO29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39mbc AUTO			
29abri AUTO30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39phcs AUTO			
30hic AUTO31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39phcs AUTO			
31mhcg AUTO32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39varantee			
32mhct AUTO33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39> AUTO			
33mhcs AUTO34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39varantee			
34mhcn AUTO35mhci AUTO36phcg AUTO37phct AUTO38phcs AUTO39v< AUTO			
35 mhci AUTO 36 phcg AUTO 37 phct AUTO 38 phcs AUTO			
36 phcg AUTO 37 phct AUTO 38 phcs AUTO			
38 phcs AUTO		36 phog AUTO	
		38 phos AUTO	-
		+>+	

IBM SPSS Statistics Processor is ready

Unicode:ON In 83 Col 35

NUM

SPSS Syntax

- The majority of functions can be done via point + click/menu system
- In order for people to exactly replicate what you did ALWAYS paste it into syntax
- You can leave comments in syntax using '*'
- Syntax is often easier to manipulate when things need to be repeated

- SPSS needs each line of code to end with a '.' and then to run it you need to add 'execute.' after the commands
- Just highlight and hit run to do it
- SPSS will attempt to help you by showing valid commands in navy and errors in red
- Often times functions and analyses in SPSS are restricted by the 'type'
 - I only use numeric & string

SPSS

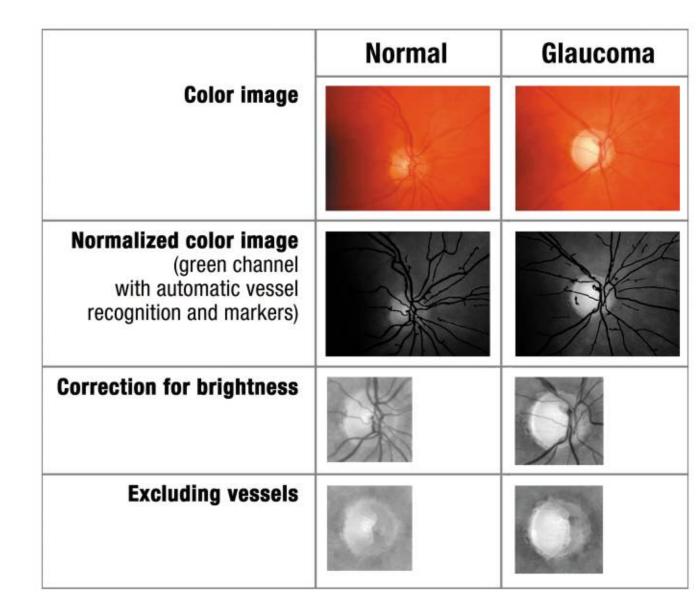
Brief Background & Example Analysis

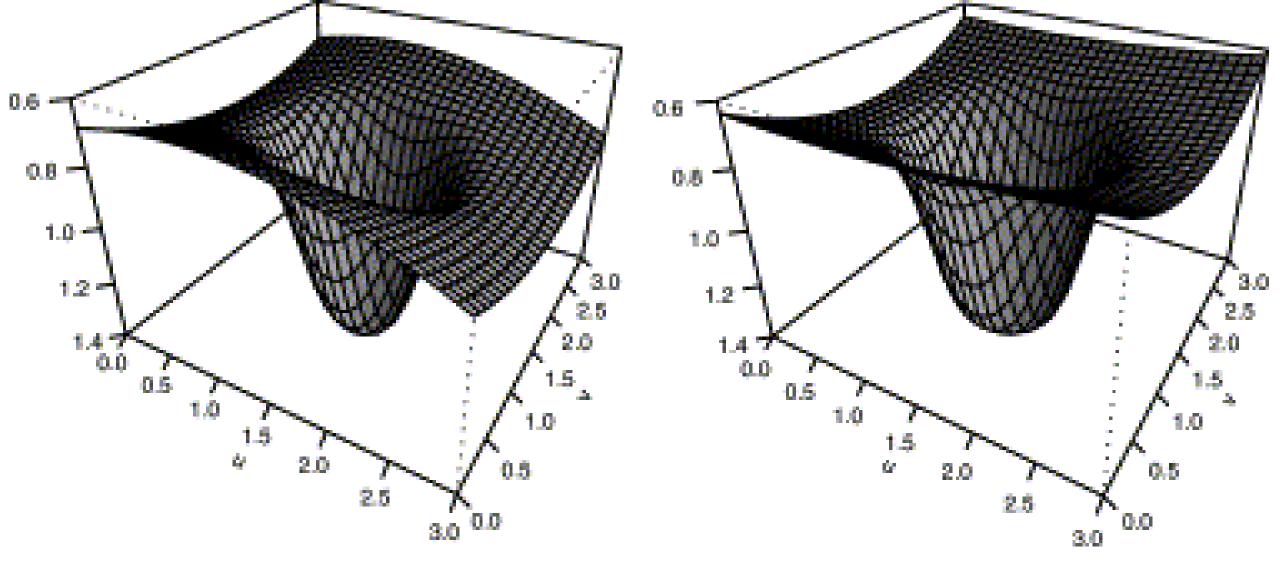
Analysis

- RQ: Predicting who might have glaucoma given measures of the optic nerve from a confocal laser scan (Heidelberg Retina Tomograph)
 - Data contains a variable with clinical diagnosis of glaucoma or normal
 - Area & Volume measures for parts of the optic nerve head measured in four locations (temporal, superior, nasal, and inferior)
 - Additionally have global measures which are sum of all four sectors
- Data & area image taken from:
 - Hothorn & Berthold Lausen (2003). Double-Bagging: Combining classifiers by bootstrap aggregation. Pattern Recognition, 36(6), 1303–1309
- Background & nerve image taken from:
 - Michelson, G., Hornegger, J., Wärntges, S., & Lausen, B. (2008). The Papilla as Screening Parameter for Early Diagnosis of Glaucoma. *Deutsches Ärzteblatt International*, 105(34-35), 583–589

Brief Background

- Roughly 67 million people suffer from glaucoma (third most common cause of blindness)
- Early diagnosis of glaucoma, is essential because by the time the patient notices functional impairment, the damage is irreversible
- Early treatment can decrease the rate of blindness 20 years later by about 50%
- As a result of the pressure, an excavation (cupping) of the optic disc may occur, along with diminution of the visual field





- Surface of the optic nerve head model for normal (left) and glaucomatous (right)
- Can sort of see that papillary excavation for the glaucomatous eye is a bit different

Analysis

- Can we predict which eye has glaucoma based on area & volume measures from the scan?
- Specifically can these three measures predict Glaucoma?
 - Effective Area in Nasal sector (EAN)
 - Peak Height Contour in Temporal (PHCT)
 - Overall Mean Radius (MR)
- Before we start, we should check to see if anyone has started working on the data set
 - I see Melissa started examining the data in R, we can open her knitted file and see what she started doing!
 - Looking through her notes, we find "Note: Per PI, subject 91 non-compliant"

In SPSS

- We will use Syntax in SPSS for data manipulation
 - Creating a new variable called status that is 0 or 1 for being normal or having glaucoma
 - We can give it a variable label
 - Can apply value labels to the coding scheme of 0 or 1