

# The Use of Survival Analysis to Compare Student Cohort Data

Dr. John Wiorkowski

Ms. Anna Moses

Dr. Lawrence J. Redlinger

Office of Strategic Planning and Analysis

Presented at the Association for Institutional Research Annual  
Forum

Orlando, Florida

May 28, 2014



THE UNIVERSITY OF TEXAS AT DALLAS

# Consortium for Student Retention Data Exchange (CSRDE)

## University of Oklahoma

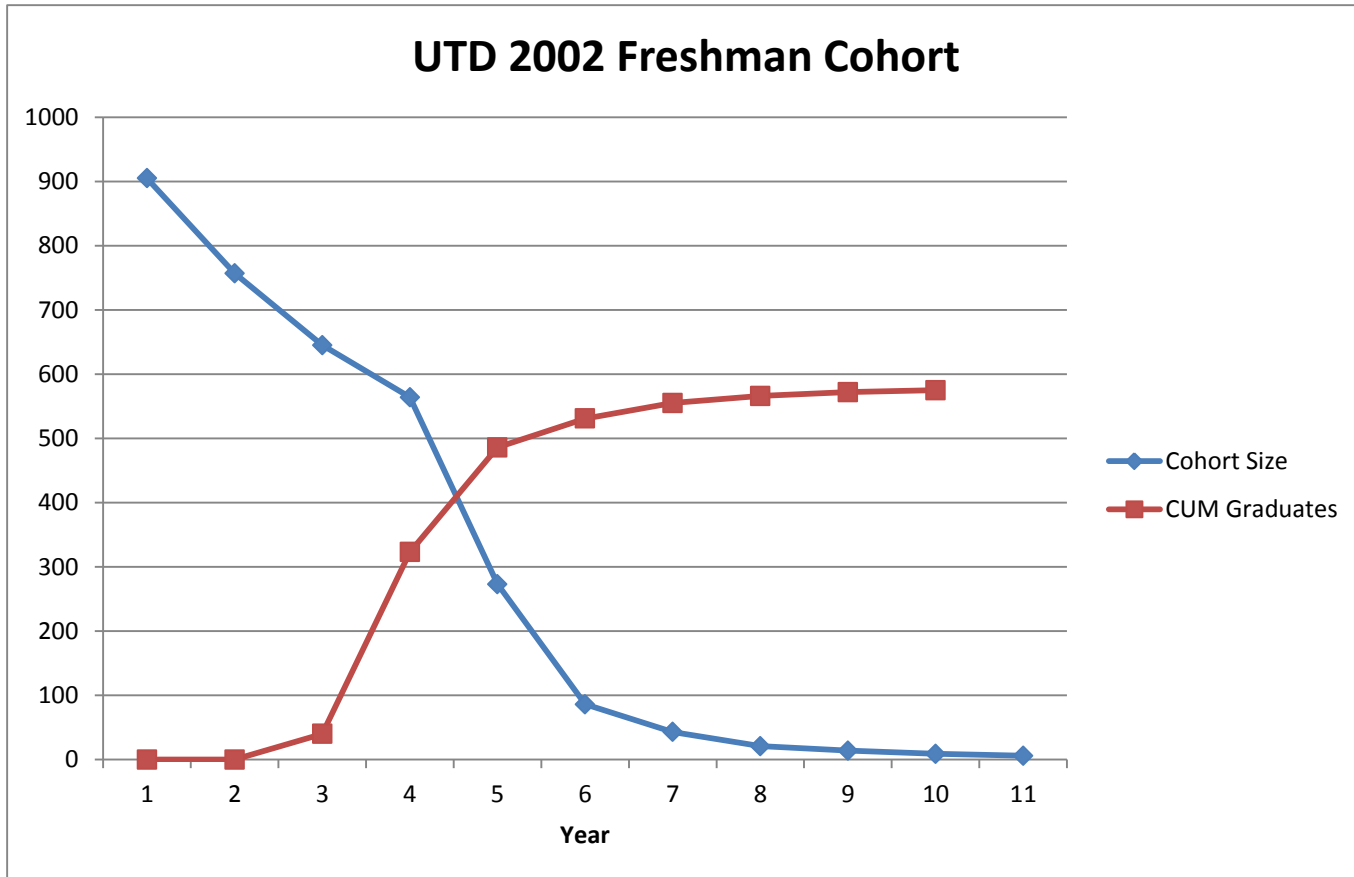
### Basic data layout

RAW DATA UT DALLAS 2002 FRESHMAN COHORT (in Percentage Terms)

Cohort Name	Cohort Size	Continued Year 2	Continued Year 3	Graduated Year 4	Continued Year 5	Graduated Year 5	Continued Year 6	Graduated Year 6	Continued Year 7	Graduated Year 7	Continued Year 8	Graduated Year 8	Continued Year 9	Graduated Year 9	Continued Year 10	Graduated Year 10	Continued Year 11
Total	905	83.6%	71.3%	35.7%	30.2%	53.7%	9.5%	58.7%	4.8%	61.3%	2.3%	62.5%	1.6%	63.2%	1.0%	63.5%	0.7%
Female	404	83.7%	71.3%	45.3%	24.0%	59.4%	6.9%	63.9%	3.0%	65.8%	1.2%	66.8%	0.7%	67.3%	0.5%	67.6%	0.5%
Male	501	83.6%	71.3%	27.9%	35.1%	49.1%	11.6%	54.5%	6.2%	57.7%	3.2%	59.1%	2.2%	59.9%	1.4%	60.3%	0.8%
Black	61	85.2%	78.7%	24.6%	42.6%	44.3%	13.1%	50.8%	9.8%	55.7%	4.9%	57.4%	0.0%	57.4%	0.0%	57.4%	1.6%
Hispanic	83	81.9%	71.1%	36.1%	26.5%	50.6%	8.4%	51.8%	8.4%	54.2%	6.0%	57.8%	2.4%	57.8%	2.4%	59.0%	0.0%
Asian	195	88.7%	77.4%	41.5%	33.3%	60.5%	11.8%	67.7%	6.7%	71.3%	3.1%	72.8%	2.1%	73.3%	2.6%	73.9%	1.5%
Anglo	539	81.8%	67.7%	34.7%	28.4%	53.1%	8.3%	57.7%	3.0%	59.7%	2.0%	60.3%	1.5%	61.2%	0.4%	61.4%	0.4%



Converting the percentages to counts and filling in the missing two numbers yields the graph below



# Data Preparation and Computation of Survival Curves

	Anglo	Cum
	Continued	Graduated
Year 1	539	
Year 2	441	
Year 3	365	26
Year 4	312	187
Year 5	153	286
Year 6	45	311
Year 7	16	322
Year 8	11	325
Year 9	8	330
Year 10	2	331
Year 11	2	



	Anglo	Cum	Original			
	Continued	Graduated	Stay	Left	Grad	With
Year 1	539		539	98	0	98
Year 2	441		441	76	0	76
Year 3	365	26	365	53	26	27
Year 4	312	187	312	159	161	-2
Year 5	153	286	153	108	99	9
Year 6	45	311	45	29	25	4
Year 7	16	322	16	5	11	-6
Year 8	11	325	11	3	3	0
Year 9	8	330	8	6	5	1
Year 10	2	331	2	0	1	-1
Year 11	2		2	2		

## Final adjusted disaggregated data

	Anglo	Cum		Original				Adjusted	Anglo			
	Continued	Graduated		Stay	Left	Grad	With	Stay	Left	Grad	With	
Year 1	539			539	98	0	98	539	89	0	89	
Year 2	441			441	76	0	76	450	76	0	76	
Year 3	365	26		365	53	26	27	374	53	26	27	
Year 4	312	187		312	159	161	-2	321	161	161	0	
Year 5	153	286		153	108	99	9	160	108	99	9	
Year 6	45	311		45	29	25	4	52	29	25	4	
Year 7	16	322		16	5	11	-6	23	11	11	0	
Year 8	11	325		11	3	3	0	12	3	3	0	
Year 9	8	330		8	6	5	1	9	6	5	1	
Year 10	2	331		2	0	1	-1	3	1	1	0	
Year 11	2			2	2			2	2			

# Computing the probability of survival for each year

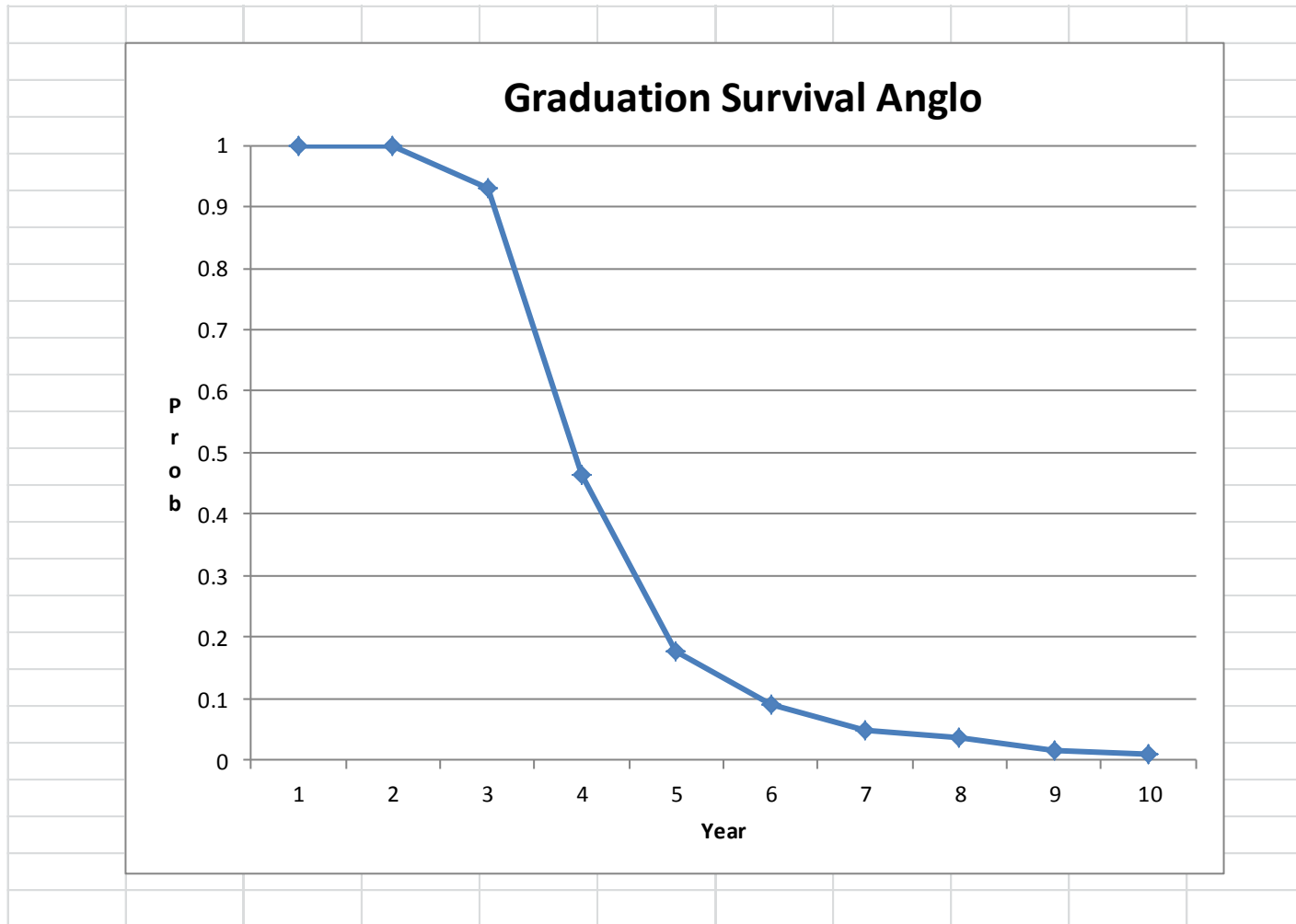
$$PROB(\text{Survive First Year}) = 1 - \frac{\text{number who graduate in first year}}{\text{number who enter first year}}$$

$$PROB(\text{Survive Second Year} \mid \text{Survive First Year}) = 1 - \frac{\text{number who graduate in second year}}{\text{number who enter second year}}$$

$$PROB(\text{Survive till Second Year}) = PROB(\text{Survive Second Year} \mid \text{Survive First Year}) PROB(\text{Survive First Year})$$

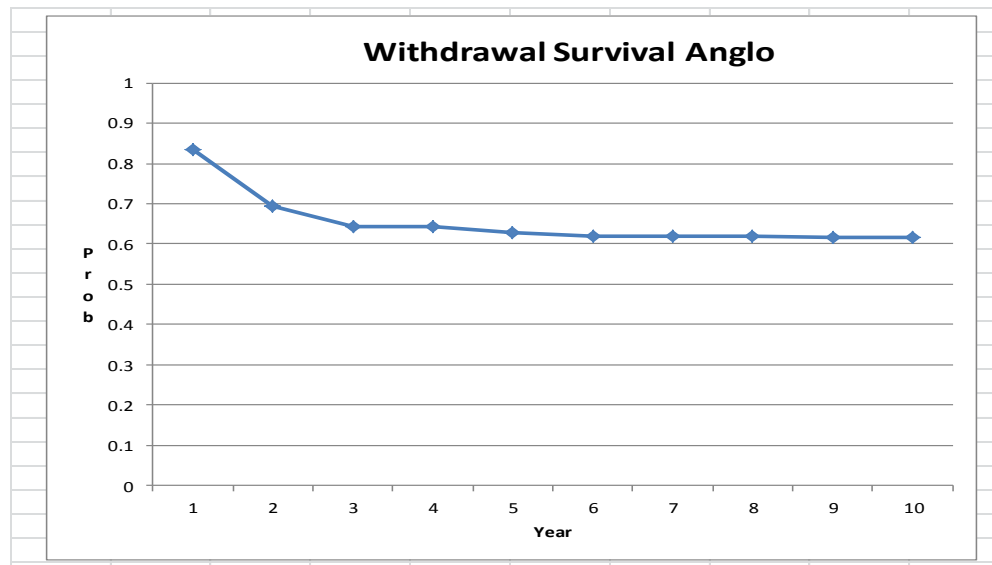
$$PROB(\text{Survive till Year } t+1) = PROB(\text{Survive Year } t+1 \mid \text{Survive till Year } t) PROB(\text{Survive till Year } t)$$

The Probability of Survival can be computed for all years and results in the following survival curve for Anglos:



# Computation of the Withdrawal Survival Curve

Original Adjusted Stay	Anglo Left	Grad	With		New Adjusted Stay	With
539	89	0	89		539	89
450	76	0	76		450	76
374	53	26	27		374	27
321	161	161	0		347	0
160	108	99	9		347	9
52	29	25	4		338	4
23	11	11	0		334	0
12	3	3	0		334	0
9	6	5	1		334	1
3	1	1	0		333	0

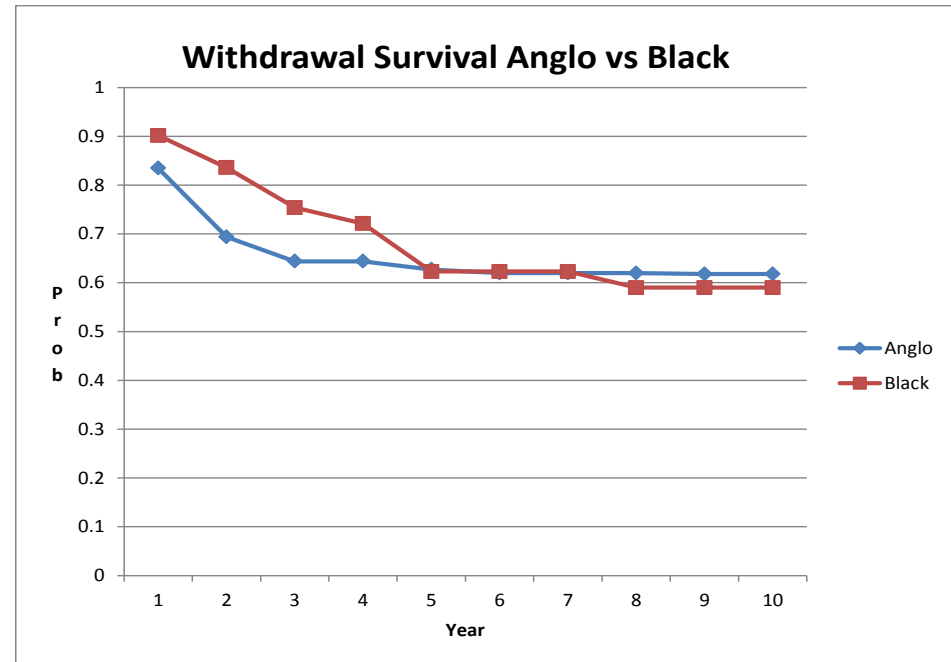
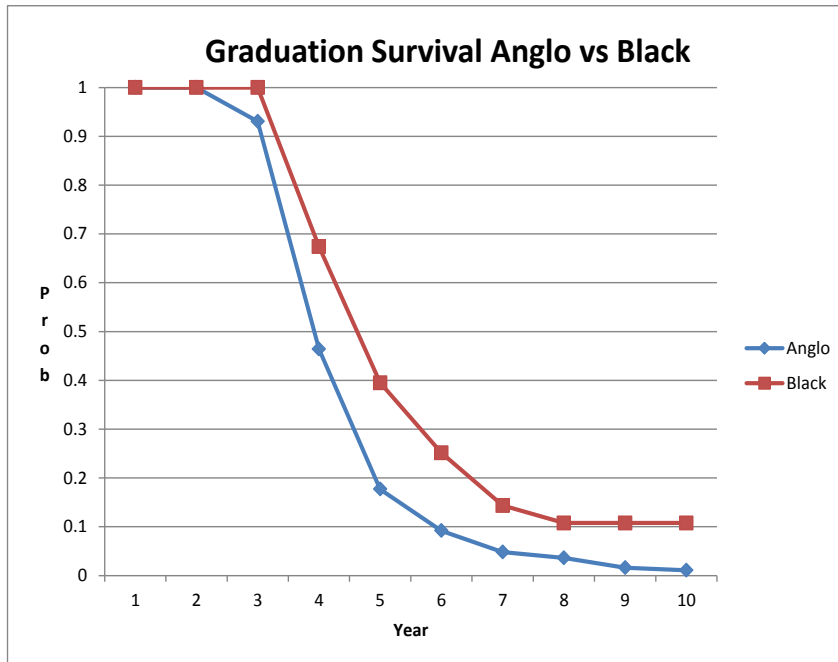




To compare Anglo and Black (for example) survival one needs to first disaggregate the Black data for the UTD 2002 cohort and adjust the data for stop-outs (Shown below).

	Black	Cum		Original	Black			Adjusted	Black		
	Continued	Graduated		Stay	Left	Grad	With	Stay	Left	Grad	With
Year 1	61			61	9	0	9	61	6	0	6
Year 2	52			52	4	0	4	55	4	0	4
Year 3	48	0		48	5	0	5	51	5	0	5
Year 4	43	15		43	17	15	2	46	17	15	2
Year 5	26	27		26	18	12	6	29	18	12	6
Year 6	8	31		8	2	4	-2	11	4	4	0
Year 7	6	34		6	3	3	0	7	3	3	0
Year 8	3	35		3	3	1	2	4	3	1	2
Year 9	0	35		0	0	0	0	1	0	0	0
Year 10	0	35		0	-1	0	-1	1	0	0	0
Year 11	1			1	1			1			

# Comparing Graduation and Withdrawal Survival



**As can be seen, there is a strong suggestion that the Anglo and Black Graduation Survival curves differ, and also that there may also be differences in the Withdrawal Survival curves of the two ethnicities. We need to formalize these observations through the use of an appropriate statistical methodology.**

## Testing Whether the Survival Curves are the Same

Year	Anglo Continue	Black Continue	Total Continue	Anglo Grad	Black Grad	Total Grad	Expected Anglo	Anglo Diff Obs - Exp	Variance	z-score
j	$N_{1j}$	$N_{2j}$	$N_j$	$O_{1j}$	$O_{2j}$	$O_j$	$E_{1j}$	$O_{1j} - E_{1j}$	$V_j$	$\frac{(O_{1j} - E_{1j})}{\sqrt{V_j}}$
1	539	61	600	0	0	0	0	0	0	NA
2	450	55	505	0	0	0	0	0	0	NA
3	374	51	425	26	0	26	22.88	3.12	2.583713	1.94
4	321	46	367	161	15	176	153.9401	7.059946	10.06922	2.22
5	160	29	189	99	12	111	93.96825	5.031746	5.982106	2.06
6	52	11	63	25	4	29	23.93651	1.063492	2.291924	0.70
7	23	7	30	11	3	14	10.73333	0.266667	1.381762	0.23
8	12	4	16	3	1	4	3	0	0.6	0.00
9	9	1	10	5	0	5	4.5	0.5	0.25	1.00
10	3	1	4	1	0	1	0.75	0.25	0.1875	0.58
							Sum	17.29185	23.34622	
	$E_{1j} = \frac{O_j}{N_j} N_{1j}$		$LR\ Test = \frac{(\sum_j (O_{1j} - E_{1j}))^2}{\sum_j V_j}$				$V_j = \frac{O_j(N_{1j}/N_j)(1 - N_{1j}/N_j)(N_j - O_j)}{N_j - 1}$			
	Chi-square	p-value								
	LR Test	12.80756	0.000345	reject hypothesis of no difference						

In order to see how they are different one can capitalize on the fact that the z-score column is a normal approximation to the exact Fisher test.

Year	Anglo Continue	Black Continue	Total Continue	Anglo With	Black With	Total With	Expected Anglo	Anglo Diff Obs - Exp	Variance	z-score	
j	$N_{1j}$	$N_{2j}$	$N_j$	$O_{1j}$	$O_{2j}$	$O_j$	$E_{1j}$	$O_{1j} - E_{1j}$	$V_j$	$\frac{(O_{1j} - E_{1j})}{\sqrt{V_j}}$	
1	539	61	600	89	6	95	85.34167	3.658333	7.31483	1.35	
2	450	55	505	76	4	80	71.28713	4.712871	6.546977	1.84	ignore
3	374	51	425	27	5	32	28.16	-1.16	3.132136	-0.66	
4	347	46	393	0	2	2	1.765903	-1.7659	0.206169	-3.89	ignore
5	347	44	391	9	6	15	13.31202	-4.31202	1.444253	-3.59	ignore
6	338	38	376	4	0	4	3.595745	0.404255	0.360493	0.67	
7	334	38	372	0	0	0	0	0	0	NA	
8	334	38	372	0	2	2	1.795699	-1.7957	0.182937	-4.20	ignore
9	334	36	370	1	0	1	0.902703	0.097297	0.087831	0.33	
10	333	36	369	0	0	0	0	0	0	NA	
							Sum	-0.16087	19.27562		

$E_{1j} = \frac{O_j}{N_j} N_{1j}$	$LR \text{ Test} = \frac{(\sum_j (O_{1j} - E_{1j}))^2}{\sum_j V_j}$	$V_j = \frac{O_j(N_{1j}/N_j)(1 - N_{1j}/N_j)(N_j - O_j)}{N_j - 1}$
-----------------------------------	---	--

			p-value	
LR Test	0.001343	0.970772		Accept the null hypothesis

# Selected Summary of Results

## **2002 UTD Freshman Cohort statistically significant results:**

- 1. Females graduated at higher rates than males in the 3<sup>rd</sup> and 4<sup>th</sup> years***
- 2. Males withdrew at higher rates than females in the 3<sup>rd</sup> and 4<sup>th</sup> years***
- 3. Anglos graduated at higher rates than Blacks in the 3<sup>rd</sup>, 4<sup>th</sup> and 5<sup>th</sup> years***
- 4. Asians graduated at a higher rate than Blacks in the 4<sup>th</sup> year***
- 5. Hispanics graduated at a higher rate than Blacks in the 5<sup>th</sup> year***
- 6. Anglos withdrew at a higher rate than Asians in the 1<sup>st</sup> and 3<sup>rd</sup> years but at a lower rate in the 4<sup>th</sup> year***
- 7. Blacks withdrew at a higher rate than Asians in the 3<sup>rd</sup> and 5<sup>th</sup> years***
- 8. Hispanics withdrew at a higher rate than Asians in the 1<sup>st</sup> and 3<sup>rd</sup> years***

## Series 1 (Females) and Series 2 (Males) data in Excel –read into SAS

	A	B	C	D	E	F	G
1	R1	R2	Year				
2	404	0	1				
3	338	0	2				
4	288	27	3				
5	252	183	4				
6	97	240	5				
7	28	258	6				
8	12	266	7				
9	5	270	8				
10	3	272	9				
11	2	273	10				
12	2	0	11				
13							
14							
15							

	A	B	C	D	E	F	G
1	R3	R4	Year				
2	501	0	1				
3	419	0	2				
4	357	13	3				
5	312	140	4				
6	176	246	5				
7	58	273	6				
8	31	289	7				
9	16	296	8				
10	11	300	9				
11	7	302	10				
12	4	0	11				
13							
14							
15							

## SAS Output – Original Data (Females)

### Series 1 Information

#### Female

Year	Cum Stay	Cum Grad
1	404	0
2	338	0
3	288	27
4	252	183
5	97	240
6	28	258
7	12	266
8	5	270
9	3	272
10	2	273
11	2	0

# SAS Output Disaggregated Data

## Diaggregated Matrix

### Female

Year	Stay	Left	Grad	With
1	404	66	0	66
2	338	50	0	50
3	288	36	27	9
4	252	155	156	-1
5	97	69	57	12
6	28	16	18	-2
7	12	7	8	-1
8	5	2	4	-2
9	3	1	2	-1
10	2	0	1	-1
11	2	0	0	0

## Disaggregated Matrix Adjusted for Stop Outs

### Female

Year	Stay	Left	Grad	With
1	404	58	0	58
2	346	50	0	50
3	296	36	27	9
4	260	156	156	0
5	104	69	57	12
6	35	18	18	0
7	17	8	8	0
8	9	4	4	0
9	5	2	2	0
10	3	1	1	0
11	2	0	0	0



# Outputs the graduation and withdrawal data for analysis

## Final Series for Analysis. Graduation Curves.

### Female and Male

Year	Female_Survival	Female_Graduates	Male_Survival	Male_Graduates
1	404	0	501	0
2	346	0	422	0
3	296	27	360	13
4	260	156	315	127
5	104	57	179	106
6	35	18	61	27
7	17	8	34	16
8	9	4	18	7
9	5	2	11	4
10	3	1	7	2
11	2	0	4	0

## Final Series for Analysis. Withdrawal Curves.

### Female and Male

Year	Female_Survival	Female_Withdraws	Male_Survival	Male_Withdraws
1	404	58	501	79
2	346	50	422	62
3	296	9	360	32
4	287	0	328	9
5	287	12	319	12
6	275	0	307	0
7	275	0	307	0
8	275	0	307	0
9	275	0	307	0
10	275	0	307	1
11	275	0	306	0

# SAS Output: Computation of the Log Rank Test for the Graduation Survival Curve

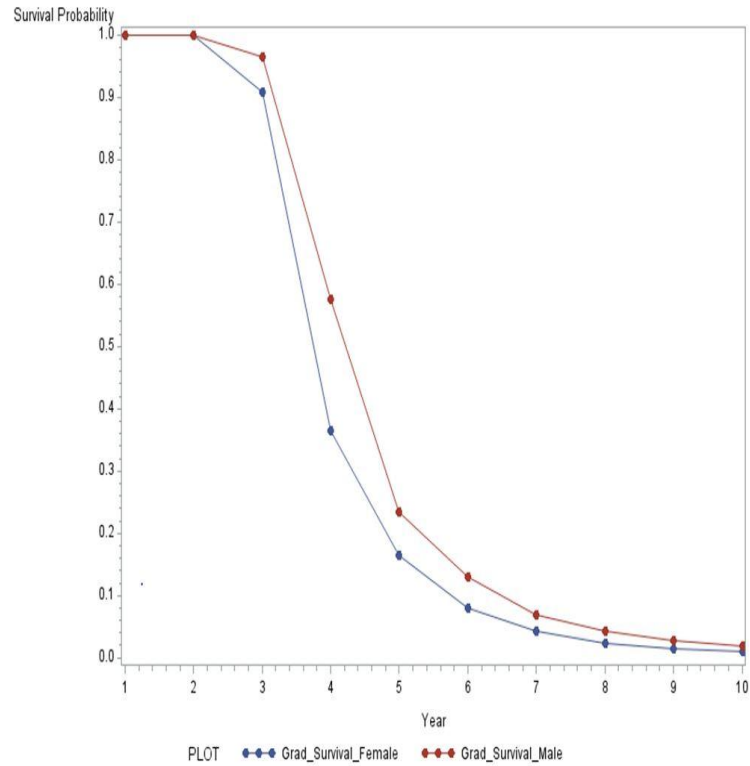
Test of Equality of Graduation Survival Curves.

Year	Obs	Exp	Var	Zscore
1	0.00	00.00	00.00	00.00
2	0.00	00.00	00.00	00.00
3	27.00	18.05	09.32	02.93
4	156.0	128.0	35.66	04.69
5	57.00	59.90	16.12	-0.72
6	18.00	16.41	05.60	00.67
7	8.00	08.00	02.88	00.00
8	4.00	03.67	01.50	00.27
9	2.00	01.88	00.86	00.13
10	1.00	00.90	00.49	00.14

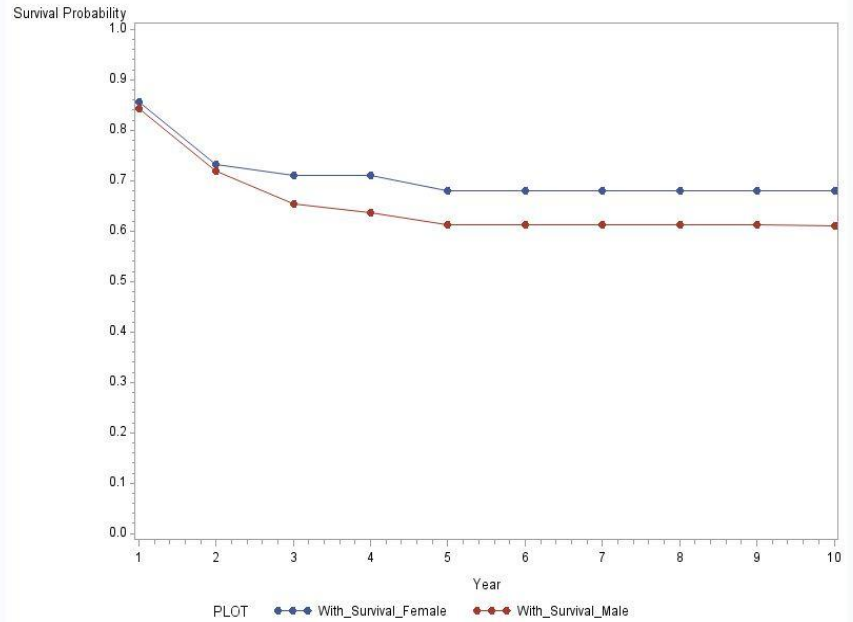
**Log Rank Test**  
**18.129604**  
**Pvalue**  
**0.0000206**

# Output plots Graduation Survival & Withdrawal Curves

## Graduation Survival

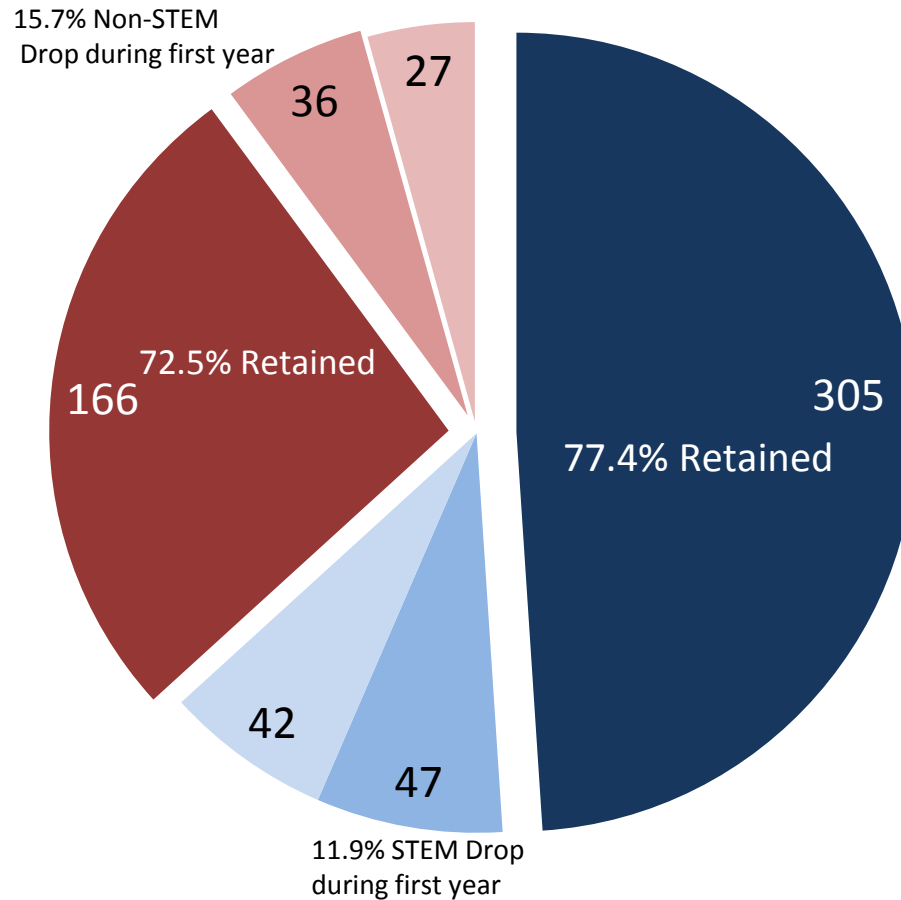


## Withdrawal Survival



# Derivative Analysis Examples

## Males



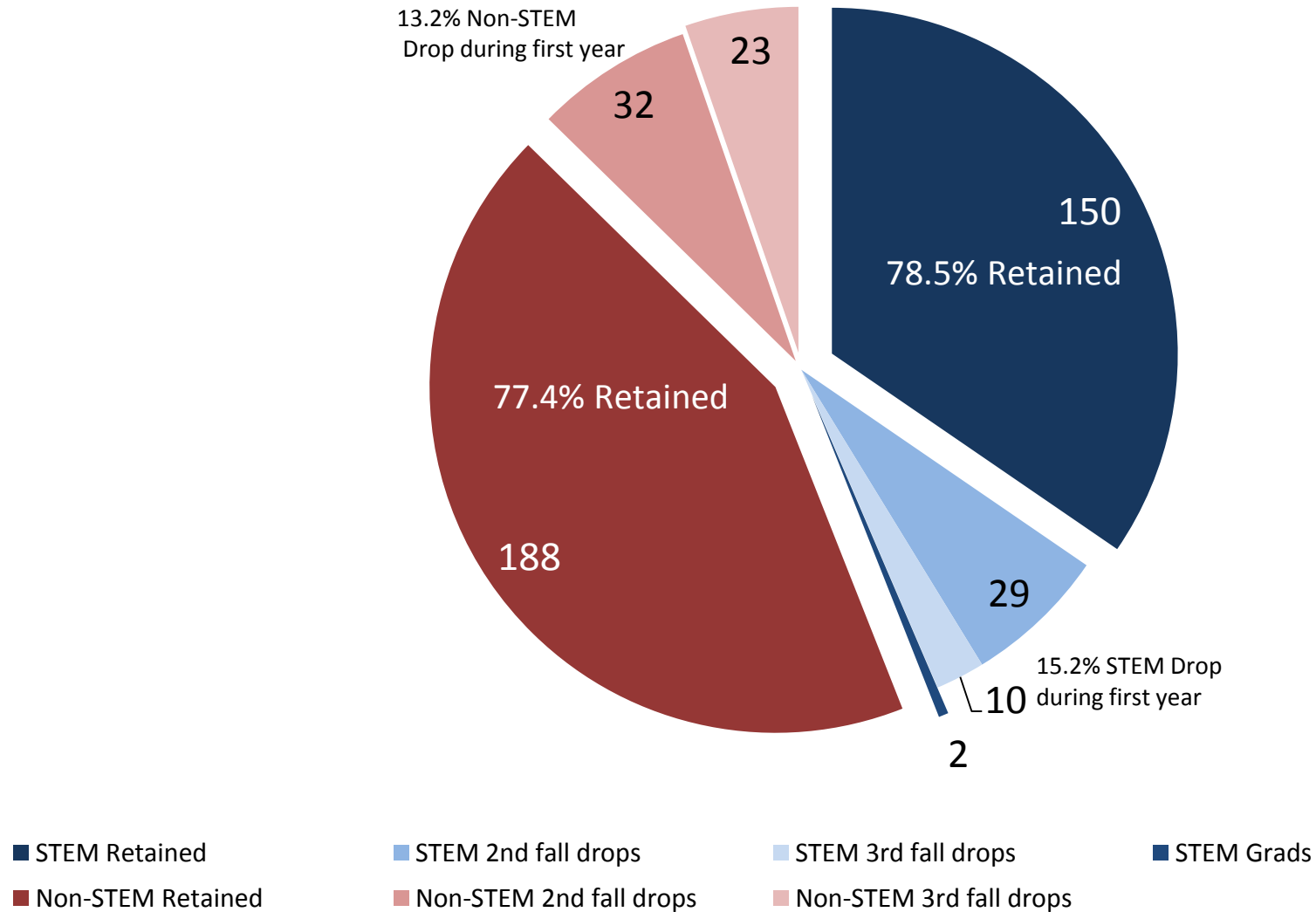
■ STEM Retained  
■ Non-STEM Retained

■ STEM 2nd fall drops  
■ Non-STEM 2nd fall drops

■ STEM 3rd fall drops  
■ Non-STEM 3rd fall drops

# Derivative Analysis Examples

## Females



# Derivative Analysis Examples

FTIC Fall 2007 Cohort - Males in Selected STEM Majors							
Beginning STEM Major	Not Enrolled	Enrolled*	Total	2nd Fall Drop Out Percent	2nd to 3rd Fall Drop Out Percent	Cum. Drop Out By 3rd Fall Percent	
Computer Engineering	15	52	67	10.4%	11.9%	22.4%	
Computer Science	27	76	103	13.6%	12.6%	26.2%	
Electrical Engineering	14	56	70	14.3%	5.7%	20.0%	
Neuroscience	3	16	19	10.5%	5.3%	15.8%	
Biology	13	36	49	18.4%	8.2%	26.5%	

FTIC Fall 2007 Cohort - Females in Selected STEM Majors							
Beginning STEM Major	Not Enrolled	Enrolled*	Total	2nd Fall Drop Out Percent	2nd to 3rd Fall Drop Out Percent	Cum. Drop Out By 3rd Fall Percent	
Computer Engineering	1	5	6	16.7%	0.0%	16.7%	
Computer Science	2	9	11	9.1%	9.1%	18.2%	
Electrical Engineering	2	11	13	7.7%	7.7%	15.4%	
Neuroscience	2	23	25	4.0%	4.0%	8.0%	
Biology	22	55	77	23.4%	5.2%	28.6%	

Male-Female variations in survival by selected majors

# Derivative Analysis Examples

Graduation in 6 Years: Gender by 1st Semester GPA									
Not Graduated in 6 years (N = 387)									
	0-0.49	0.5-0.99	1-1.49	1.5-1.99	2-2.49	2.5-2.99	3-3.49	3.5-4	Total
Female	6	2	12	8	24	24	30	23	129
Male	29	15	13	30	46	47	57	21	258
Total	35	17	25	38	70	71	87	44	387
Graduated in 6 years (N = 670)									
	0-0.49	0.5-0.99	1-1.49	1.5-1.99	2-2.49	2.5-2.99	3-3.49	3.5-4	Total
Female	0	0	0	1	14	33	98	159	305
Male	1	1	2	7	26	54	134	140	365
Total	1	1	2	8	40	87	232	299	670
Graduation Rate									
	0-0.49	0.5-0.99	1-1.49	1.5-1.99	2-2.49	2.5-2.99	3-3.49	3.5-4	Total
Female	0.00%	0.00%	0.00%	11.11%	36.84%	57.89%	76.56%	87.36%	70.28%
Male	3.33%	6.25%	13.33%	18.92%	36.11%	53.47%	70.16%	86.96%	58.59%
Total	2.78%	5.56%	7.41%	17.39%	36.36%	55.06%	72.73%	87.17%	63.39%

- 29 or 6.7% of females had first semester GPAs below 2.0
- 98 or 15% of males had first semester GPAs below 2.0
- 257 or 59% of females had first semester GPAs above a 3.0
- 274 or 44% of males had first semester GPAs above a 3.0

Thank you for your time.

Now we would be happy to answer your questions.

Contacts:

Ms. Anna Moses

Assistant Director

Office of Strategic Planning and Analysis

The University of Texas at Dallas

Phone: (972) 883-4188

[Anna.moses@utdallas.edu](mailto:Anna.moses@utdallas.edu)