Session 184: Survival Analysis
Annual Meeting 2015

John McGarry PhD, ASA
jmcgarry@insightdecision.com
https://ca.linkedin.com/in/johnkmcgarry
www.insightdecision.com

Insight Decision Solutions Inc.

Agenda

- Introduction
- Methods
- Survival Time
- Standard Error
- Risk Factors

- Updated Presentation
  - http://insightdecision.com/articles/

Insight Decision Solutions Inc.

Introduction

- Survival analysis quantifies the risk of decrement from a population for one or more events
  - Hazard ratio, relative risk to base
  - Survival time, average time to event
  - Survival rate, probability of survival to time *t*
- Used across multiple disciplines
  - Survival – Clinical Medicine
  - Reliability/Failure - Engineering
  - Event History – Sociology
  - Duration – Economics
- Insurance
  - Persistency
  - Claim Continuation

Insight Decision Solutions Inc.

Examples

- Clinical, Actuarial
  - Time to death
  - Time to incidence
  - Time to recovery or relapse
- Engineering
  - Time to failure
- Operational
  - Time to new business decline
  - Time to new business issue
  - Time to new business complete
  - Time to claim decline
  - Time to claim settlement
  - Time to claim closure

Insight Decision Solutions Inc.

Implicit Concepts 1

- State
  - Cancer Diagnosis - Clinical
  - Insured Life – Insurance
  - Operating - Engineering
  - Application in Underwriting – Insurance
  - Claim in Process - Insurance
- Event
  - Cancer Death
  - Death, Lapse
  - Failure
  - Issue, Decline, Complete
  - Settlement, Decline, Closure

Implicit Concepts 2

- Time Origin
  - Entry into State
- Time Scale
  - Days
  - Months
  - Quarters
  - Years

Insight Decision Solutions Inc.

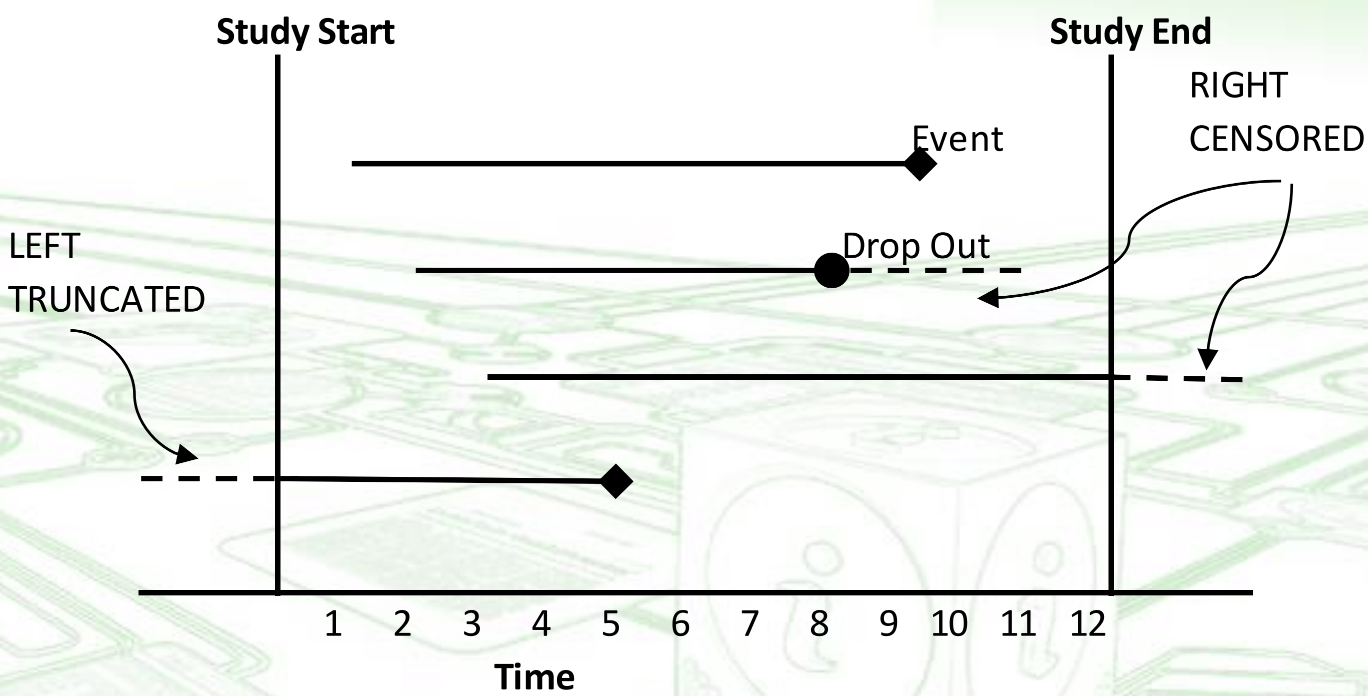
Other Events

- Exit from Population due to events other than specified event – Right Censored
  - Cancer: drop outs, non-cancer deaths
  - Mortality: lapse
  - Lapse: death
- Independent or non-informative if they are not related to the study event
- Dependent or informative if related, e.g.
  - Sick patients drop out of cancer study
  - Healthy lives lapse policies
- Censored events assumed independent

Study Period

- Study defined by
  - Calendar period: Start Date to End Date
    - 1st Jan 2013 – 31st Dec 2015
  - Duration: Period from entry
    - 1st Year/12 Months in State
- Right Censored
  - Period after study end, non-informative
- Left Truncated
  - For lives entering state before study start
- Right Truncated
  - Only lives exiting state due to decrement before study end

Summary Graphic



Insight Decision Solutions Inc.

Methods

- Simple
- Actuarial
- Kaplan Meier

Simple Method

- No truncation or censoring
  - All lives enter during study period
  - No drop outs/withdrawals
  - Study ends when last life suffers decrement
- Probability of decrement before t_i
  - ${}_iq_0 = {}_iD_0/N$
- ${}_iD_0$= Decrements before time t_i
- N = Number of lives in study
- Survival rate, i.e. the probability of surviving to time t_i
  - $S(t_i) = {}_ip_0 = (1 - {}_iD_0/N)$

Insight Decision Solutions Inc.

Actuarial Method 1

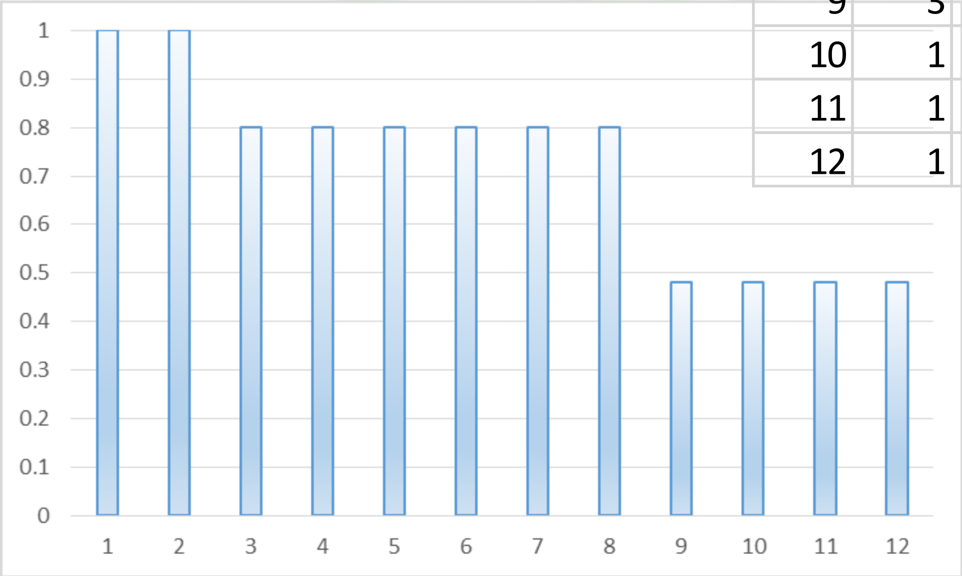
- Exposure defined over durational period divided into equal time periods: $t_j; j = 1, n$
- Probability of decrement in current period t_j
  - $q_j = D_j / E_j$
- D_j = Decrements over period t_j
- $E_j = N_j - 0.5W_j$ = Exposure over period t_j
  - Independent or single decrement rate
  - $E_j = N_j$ for dependent, multiple decrement rate
- N_j= Number of lives at start of period t_j
- W_j = Withdrawals over period t_j
- Force, instantaneous failure rate, hazard - μ, λ
  - $h_{j+0.5} = D_j / (E_j - 0.5D_j)$, $h_j = (D_{j-1} + D_j)/(2N_j)$

Actuarial Method 2

- Probability of surviving the current period, or conditional survival given survival to end of prior period, $p_j = {}_1p_j$
  - $p_j = 1 - q_j = 1 - D_j/E_j$
- Survival rate, i.e. the probability of surviving past time t_i from entry.
  - $S(t_i) = {}_ip_0 = \prod_j (1 - D_j/E_j)$
    - $S(t_1) = (1 - D_1/E_1)$
    - $S(t_2) = (1 - D_1/E_1)(1 - D_2/E_2)$
- Life Table or Actuarial Estimator

Actuarial Example

t	Nt	Dt	Wt	Et	St
1	5			5.00	1.00
2	5			5.00	1.00
3	5	1		5.00	0.80
4	4			4.00	0.80
5	4			4.00	0.80
6	4		1	3.50	0.80
7	3			3.00	0.80
8	3			3.00	0.80
9	3	1	1	2.50	0.48
10	1			1.00	0.48
11	1			1.00	0.48
12	1			1.00	0.48



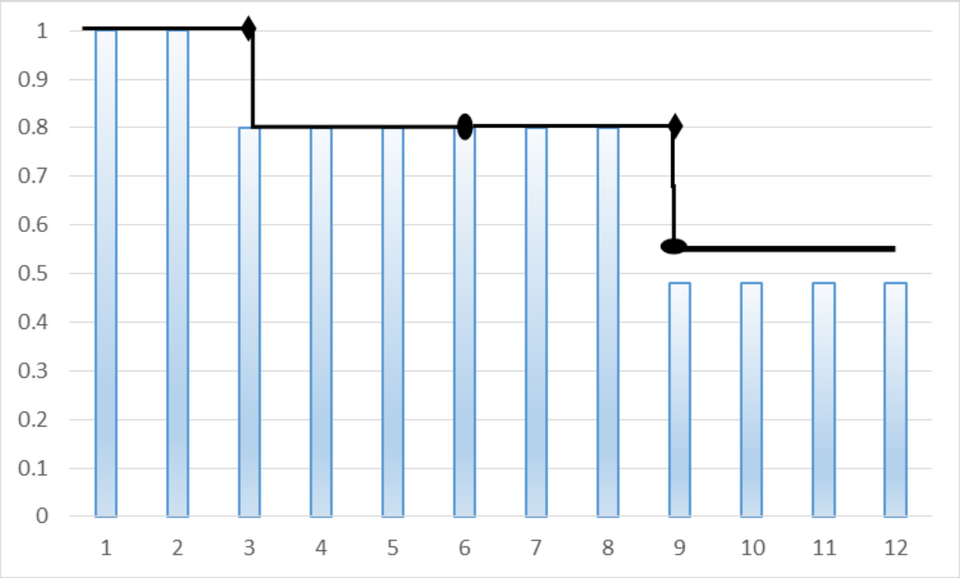
Kaplan Meier Method

- Take increasingly small intervals of time - Product Limit Estimator
- Each event occurs at time $t_k; k = 1, m$, giving irregular time intervals, (t_{k-1}, t_k) and for tied events in t_k, assume withdrawals last
- Survival rate
  - $S(t) = \prod_{k:t_k<t}(1 - D_k/E_k)$
- D_k= number of decrements at time t_k
- $E_k = N_k$= lives at risk at time t_k
- Hazard, force of decrement, instantaneous or age specific failure rate - λ, μ
  - $h_k = D_k/E_k$
- Monthly actuarial method (dependent rates) is equivalent to monthly KM for large sample

Kaplan Meier Example

- KM curve is discontinuous, stepping down at each decrement time with censored events plotted on curve.
- Useful for small samples to see decrements, less useful for large samples.

t	Nt	Dt	Wt	St
3	5	1		0.80
6	4		1	0.80
9	3	1	1	0.53



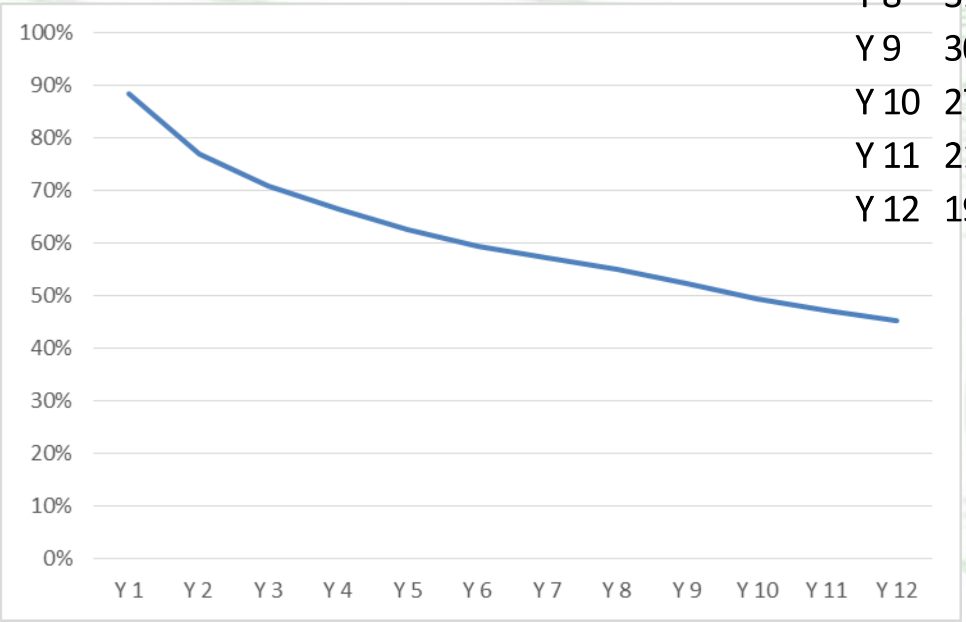
Examples

- Persistency
- Claim Closure
- Actuarial Method

Persistency Example 1

- Annual persistency rates
- 12 Years

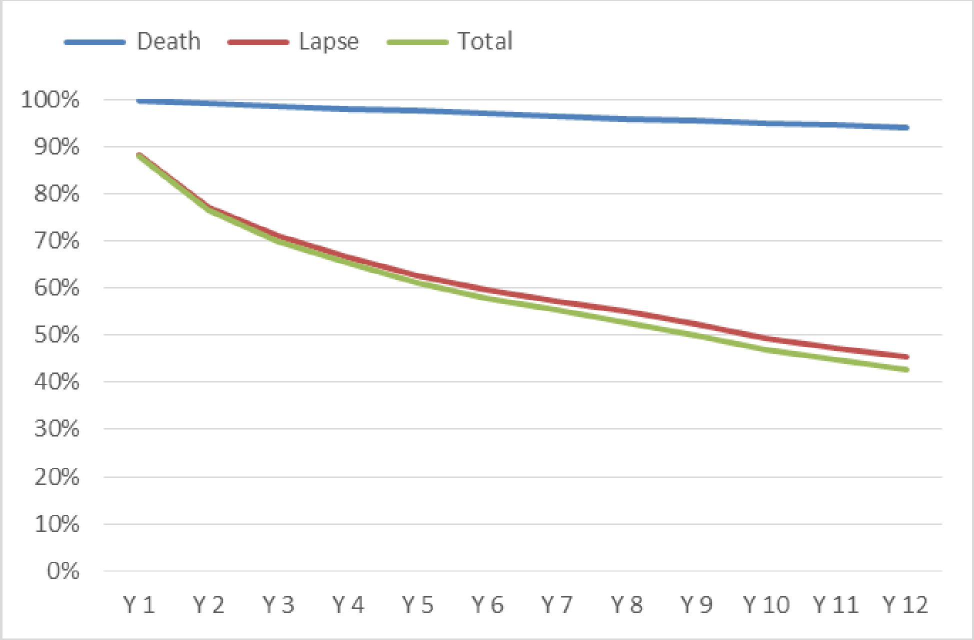
t	Et	Dt	Wt	St
Y 1	57,636	6,687	267	88.40%
Y 2	49,091	6,256	341	77.13%
Y 3	42,412	3,385	294	70.98%
Y 4	39,096	2,452	238	66.52%
Y 5	36,971	2,176	188	62.61%
Y 6	35,019	1,729	222	59.52%
Y 7	33,731	1,288	216	57.25%
Y 8	31,985	1,267	205	54.98%
Y 9	30,086	1,433	171	52.36%
Y 10	27,257	1,513	3,249	49.45%
Y 11	21,455	957	604	47.25%
Y 12	19,170	799	200	45.28%



Insight Decision Solutions Inc.

Persistency Example 2

- By Termination Reason



t	Death	Lapse	Total
Y 1	100%	88%	88%
Y 2	99%	77%	76%
Y 3	99%	71%	70%
Y 4	98%	67%	65%
Y 5	98%	63%	61%
Y 6	97%	60%	58%
Y 7	96%	57%	55%
Y 8	96%	55%	53%
Y 9	95%	52%	50%
Y 10	95%	49%	47%
Y 11	95%	47%	45%
Y 12	94%	45%	43%

Persistency Example 3

- By Issue Year

t	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
Y 1												89%	88%	88%	88%	88%	90%	88%
Y 2											90%	79%	78%	77%	74%	75%		77%
Y 3										94%	84%	74%	71%	70%	67%			71%
Y 4									95%	89%	79%	69%	66%	65%				67%
Y 5								96%	91%	83%	74%	64%	61%					63%
Y 6							97%	92%	87%	79%	69%	60%						60%
Y 7						99%	94%	89%	83%	75%	66%							57%
Y 8					99%	96%	90%	85%	79%	69%								55%
Y 9				99%	95%	91%	85%	80%	74%									52%
Y 10			99%	95%	90%	85%	79%	74%										49%
Y 11		98%	96%	91%	85%	81%	75%											47%
Y 12	99%	95%	91%	86%	81%	76%												45%

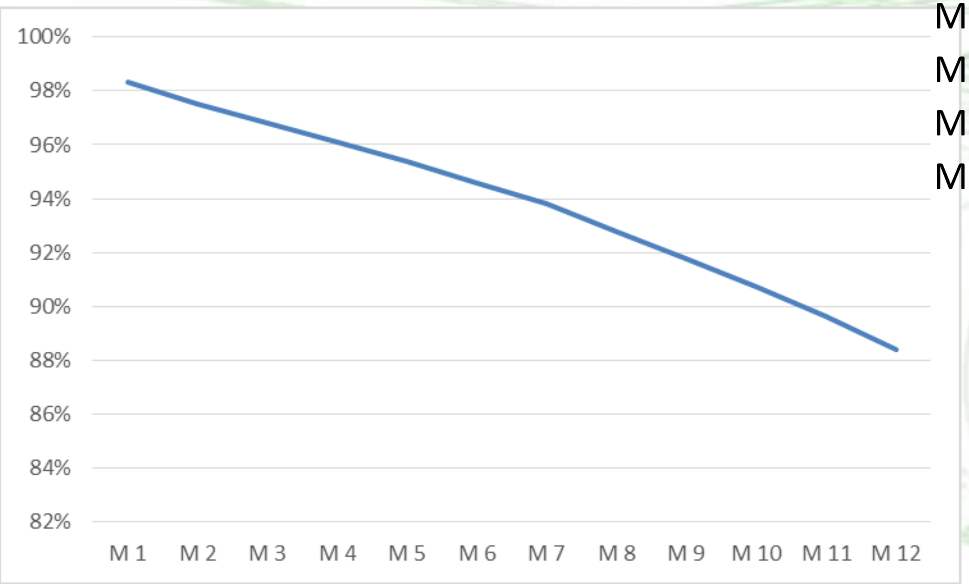
Left Truncated

Right Censored

Persistency Example 4

- Monthly persistency rates
- 1 Year

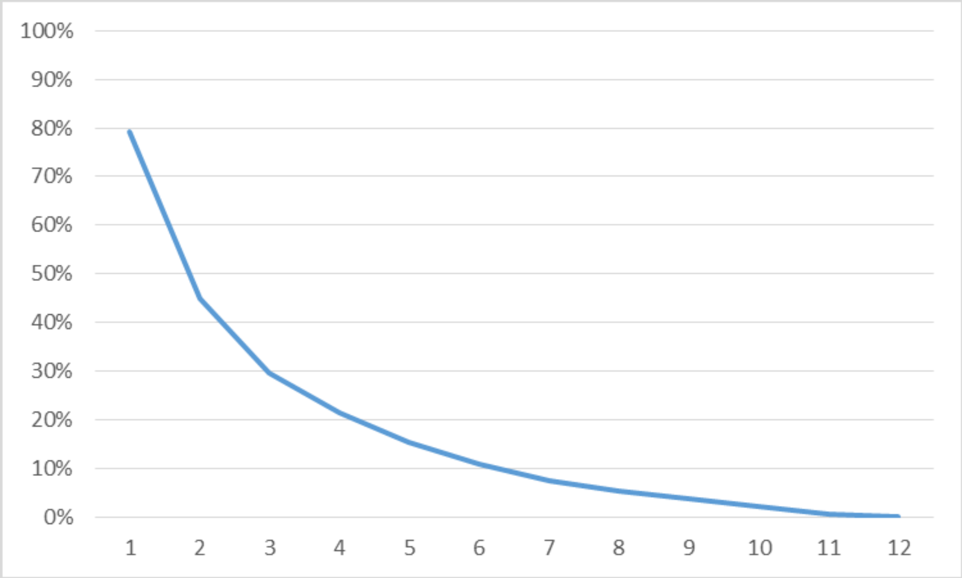
t	Et	Dt	Wt	St
M 1	58,064	964	14	98.34%
M 2	57,050	462	7	97.54%
M 3	56,563	435	15	96.79%
M 4	56,057	400	22	96.10%
M 5	55,575	429	21	95.36%
M 6	55,037	444	28	94.59%
M 7	54,457	460	23	93.79%
M 8	53,918	599	30	92.75%
M 9	53,017	569	25	91.75%
M 10	52,375	590	22	90.72%
M 11	51,722	634	33	89.61%
M 12	51,021	701	27	88.38%



Claim Closure Example 1

- Operational– turn-around time
- RBNS – pending claims with “haircut” for declines
- IBNS - stable development pattern

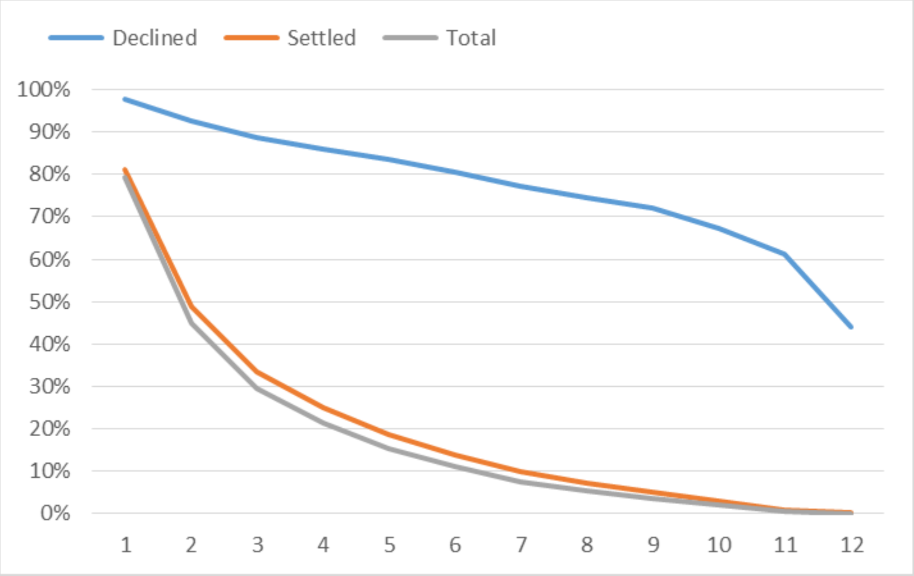
t	Et	Dt	St
1	22,549	4,665	79.31%
2	17,139	7,411	45.02%
3	9,217	3,179	29.49%
4	5,816	1,613	21.31%
5	4,061	1,118	15.44%
6	2,826	822	10.95%
7	1,911	608	7.47%
8	1,258	384	5.19%
9	855	266	3.57%
10	560	243	2.02%
11	296	218	0.53%
12	66	65	0.01%



Claim Closure Example 2

- By Closed Reason

t	Declined	Settled	Total
1	98%	81%	79%
2	93%	49%	45%
3	89%	33%	29%
4	86%	25%	21%
5	84%	19%	15%
6	81%	14%	11%
7	77%	10%	7%
8	75%	7%	5%
9	72%	5%	4%
10	67%	3%	2%
11	61%	1%	1%
12	44%	0%	0%



Insight Decision Solutions Inc.

Claim Closure Example 3

- By Opened Month
- 2015 Q3TD Monthly

Month	2015-01	2015-02	2015-03	2015-04	2015-05	2015-06	2015-07	2015-08	2015-09	2015
1	87%	84%	72%	71%	66%	65%	70%	76%	99%	76%
2	59%	41%	36%	38%	35%	32%	32%	60%		43%
3	43%	26%	22%	22%	21%	21%	28%			29%
4	30%	19%	15%	17%	16%	18%				21%
5	24%	13%	10%	11%	15%					16%
6	15%	7%	7%	11%						11%
7	8%	4%	6%							7%
8	5%	3%								5%
9	5%									4%

Probability Functions

Related Functions

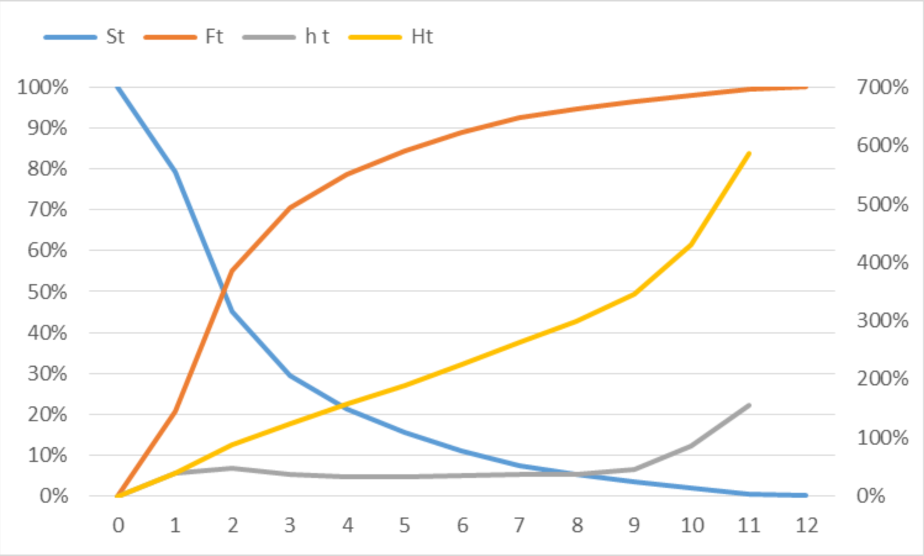
- Survival Function: Probability of surviving up to time t
  - $S(t) = Pr(T \geq t)$
- Cumulative Distribution: Probability of event occurring before time t
  - $F(t) = Pr(T < t) = 1 - S(t)$
- Density: Probability of event occurring at time t
  - $f(t) = Pr(T = t)$
- Hazard: Probability of event given survival to at time t
  - $h(t) = Pr(T = t | T \geq t) = f(t)/S(t)$
- Cumulative Hazard:
  - $H(t) = \sum_{u<t} h(u)$

$$\approx -log(S(t)) = -\sum_{u<t} log(1 - h(u))$$

(log is natural logarithm) (KM)

Insight Decision Solutions Inc.

Functions Example



t	St	Ft	h t	Ht
0	100%	0%	0%	0%
1	79%	21%	39%	39%
2	45%	55%	48%	88%
3	29%	71%	37%	124%
4	21%	79%	32%	157%
5	15%	85%	33%	190%
6	11%	89%	36%	225%
7	7%	93%	37%	262%
8	5%	95%	36%	299%
9	4%	96%	46%	345%
10	2%	98%	86%	431%
11	1%	99%	155%	586%
12	0%	100%		

Survival Time

- Mean
- Median

Mean Survival Time

- Expectation or average duration at event
  - $e = \sum_i S(t_i) = \sum_i {}_ip_0$
  - Area under survival curve
- Undiscounted annuity
  - $a = \sum_i {}_ip_0 \cdot v^i$
- With payments at times t_i then a, e are a measure of lifetime value
  - DI Claims – total cost of claim
  - Payout Annuities – total cost of benefit
  - Life – total value of premium
- May be more meaningful than average decrement rates

Mean Time Example 1

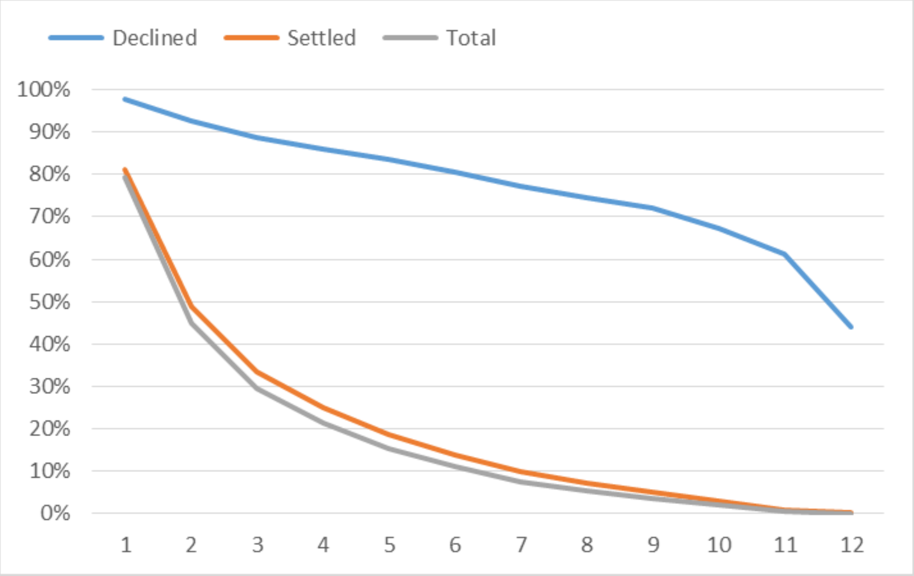
- Time to Claim Closure

t	Et	Dt	St	et
1	22,549	4,665	79.31%	0.79
2	17,139	7,411	45.02%	1.24
3	9,217	3,179	29.49%	1.54
4	5,816	1,613	21.31%	1.75
5	4,061	1,118	15.44%	1.91
6	2,826	822	10.95%	2.02
7	1,911	608	7.47%	2.09
8	1,258	384	5.19%	2.14
9	855	266	3.57%	2.18
10	560	243	2.02%	2.20
11	296	218	0.53%	2.20
12	66	65	0.01%	2.20

Mean Time Example 2

- Closure By Closed Reason

Reason	Mean t
Declined	10.14
Settled	2.47
Total	**2.20**



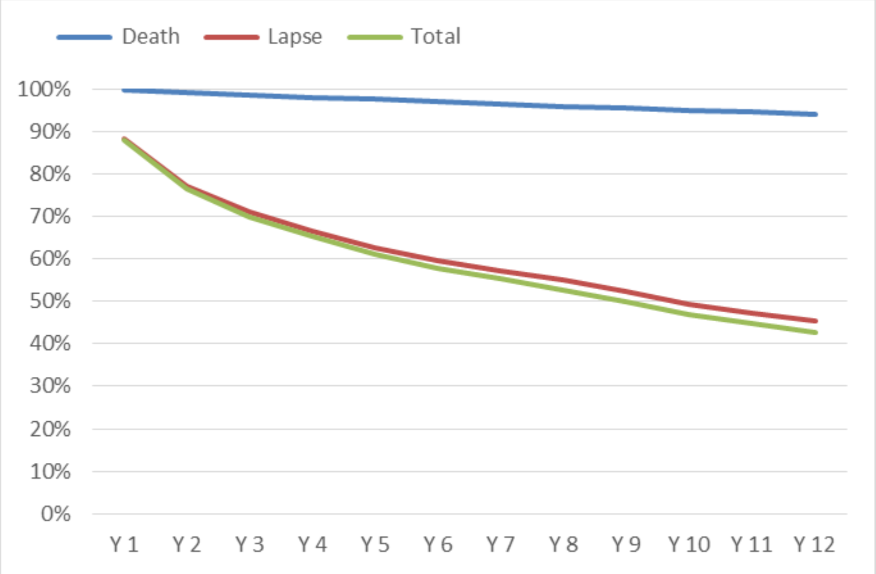
t	Declined	Settled	Total
1	98%	81%	79%
2	93%	49%	45%
3	89%	33%	29%
4	86%	25%	21%
5	84%	19%	15%
6	81%	14%	11%
7	77%	10%	7%
8	75%	7%	5%
9	72%	5%	4%
10	67%	3%	2%
11	61%	1%	1%
12	44%	0%	0%

Mean Time Example 3

- Persistency by Termination Reason

Reason	Mean t
Death	11.61
Lapse	7.32
Total	**7.11**

t	Death	Lapse	Total
Y 1	100%	88%	88%
Y 2	99%	77%	76%
Y 3	99%	71%	70%
Y 4	98%	67%	65%
Y 5	98%	63%	61%
Y 6	97%	60%	58%
Y 7	96%	57%	55%
Y 8	96%	55%	53%
Y 9	95%	52%	50%
Y 10	95%	49%	47%
Y 11	95%	47%	45%
Y 12	94%	45%	43%



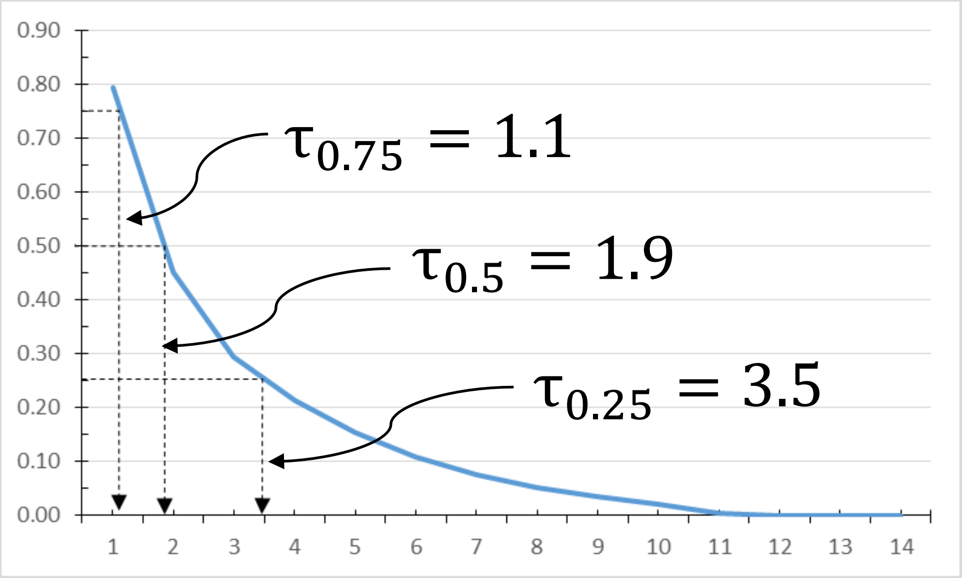
Insight Decision Solutions Inc.

Median Survival Time

- Quartiles
  - Upper Quartile $\tau_{0.75}$: $S(\tau_{0.75}) = 0.75$
  - Median $\tau_{0.5}$: $S(\tau_{0.5}) = 0.5$
  - Lower Quartile $\tau_{0.25}$: $S(\tau_{0.25}) = 0.25$
- Comments
  - Summarizes survival curve at key points
  - Where ultimate survival time not complete allows comparison at key points
  - May be more useful than mean – e.g. time at which 75% of claims are closed

Median Time Example 1

- Time to Claim Closure



Insight Decision Solutions Inc.

Median Time Example 2

- Time to Claim Closure
- 2015 Q3TD Monthly

Month	2015-01	2015-02	2015-03	2015-04	2015-05	2015-06	2015-07	2015-08	2015-09	2015
1	87%	84%	72%	71%	66%	65%	70%	76%	99%	76%
2	59%	41%	36%	38%	35%	32%	32%	60%		43%
3	43%	26%	22%	22%	21%	21%	28%			29%
4	30%	19%	15%	17%	16%	18%				21%
5	24%	13%	10%	11%	15%					16%
6	15%	7%	7%	11%						11%
7	8%	4%	6%							7%
8	5%	3%								5%
9	5%									4%
Mean	2.75	1.96	1.67	1.69	1.52	1.37	1.30	1.36	0.99	2.12
0.75	1.43	1.20	0.88	0.86	0.73	0.71	0.82	1.04	-	1.03
0.50	2.55	1.80	1.61	1.63	1.50	1.45	1.52	-	-	1.78
0.25	4.77	3.13	2.77	2.80	2.70	2.66	-	-	-	3.50

Standard Error

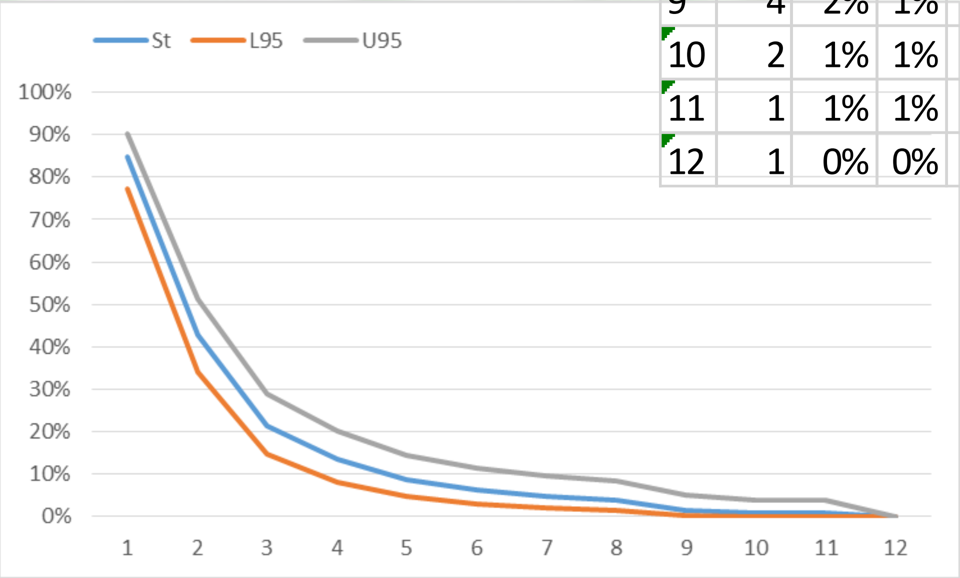
Standard Error

- Crude: $\sqrt{{}_ip_0\left(1-{}_ip_0\right)/N_1}$
- Greenwood's Formula
  - $se\left(S(t_i)\right) = S(t_i)\sqrt{\sum_j \frac{D_j}{(E_j-D_j)E_j}}$
- Raw 95% Confidence Interval
  - $S(t_i) \pm 1.96se\left(S(t_i)\right)$ - may be < 0 or > 1
- Log-log approach
  - $L(t_i) = log\left(-log\left(S(t_i)\right)\right)$
  - $se\left(L(t_i)\right) = \frac{1}{log\left(S(t_i)\right)}\sqrt{\sum_j \frac{D_j}{(E_j-D_j)E_j}}$
  - $S(t_i)e^{\pm 1.96se\left(L(t_i)\right)}$

Standard Error Example 1

- Time to Claim Closure
- Small sample
- 117 claims

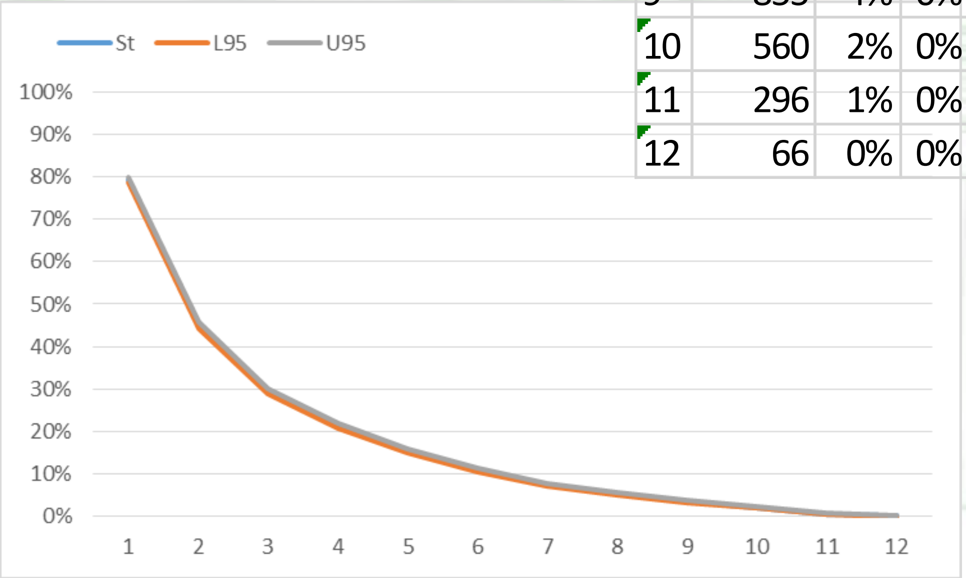
t	Et	St	set	L95	U95	CI	set/St	CI/St
1	117	85%	3%	77%	90%	13%	4%	15%
2	81	43%	4%	34%	51%	17%	10%	40%
3	41	21%	4%	15%	29%	14%	17%	66%
4	22	13%	3%	8%	20%	12%	23%	88%
5	14	9%	3%	5%	14%	10%	29%	113%
6	10	6%	2%	3%	12%	9%	34%	135%
7	7	5%	2%	2%	9%	8%	40%	158%
8	6	4%	2%	1%	8%	7%	44%	175%
9	4	2%	1%	0%	5%	5%	70%	303%
10	2	1%	1%	0%	4%	4%	100%	489%
11	1	1%	1%	0%	4%	4%	100%	489%
12	1	0%	0%	0%	0%	0%		



Standard Error Example 2

- Time to Claim Closure
- Full sample
- 22,549 claims

t	Et	St	set	L95	U95	CI	set/St	CI/St
1	22,549	79%	0%	79%	80%	1%	0%	1%
2	17,139	45%	0%	44%	46%	1%	1%	3%
3	9,217	29%	0%	29%	30%	1%	1%	4%
4	5,816	21%	0%	21%	22%	1%	1%	5%
5	4,061	15%	0%	15%	16%	1%	2%	6%
6	2,826	11%	0%	11%	11%	1%	2%	8%
7	1,911	7%	0%	7%	8%	1%	3%	10%
8	1,258	5%	0%	5%	6%	1%	3%	12%
9	855	4%	0%	3%	4%	1%	4%	15%
10	560	2%	0%	2%	2%	0%	5%	21%
11	296	1%	0%	0%	1%	0%	11%	44%
12	66	0%	0%	0%	0%	0%		



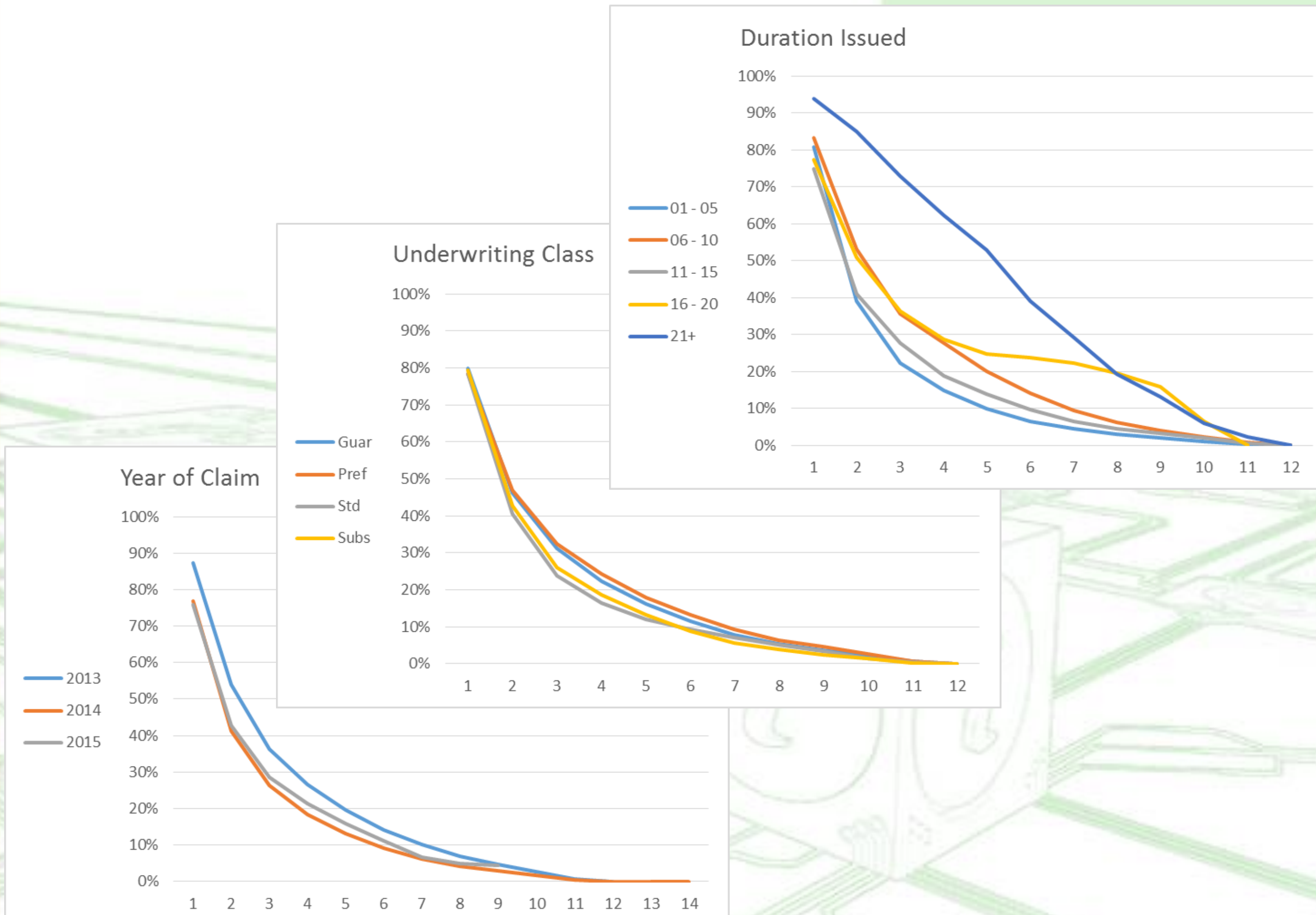
Risk Factors

- Review
- Single Factor Test - Log Rank
- Multiple Factor Test – Cox Proportional Hazards

Risk Factors

- The risk of decrement may be affected by a number of factors within the population
- Risk factors for mortality are gender, smoker status, underwriting class and occupation as well as age
- The survival behaviour for a population is based on the mix of risk factors within the population
- Comparing two different populations or the same population at different times may be misleading if there are different mixes of risk factors
- Estimating future populations requires that the mix of factors used to estimate survival behavior matches the base population to which it is applied
- Identifying risk factors requires comparing survival behaviour for each risk factor against the population
- Risk factors divide a population with heterogeneous risk into subpopulations with homogeneous risk
- Number of risk factors depends on purpose of analysis

Risk Factor Examples



Insight Decision Solutions Inc.

Single Factor Test

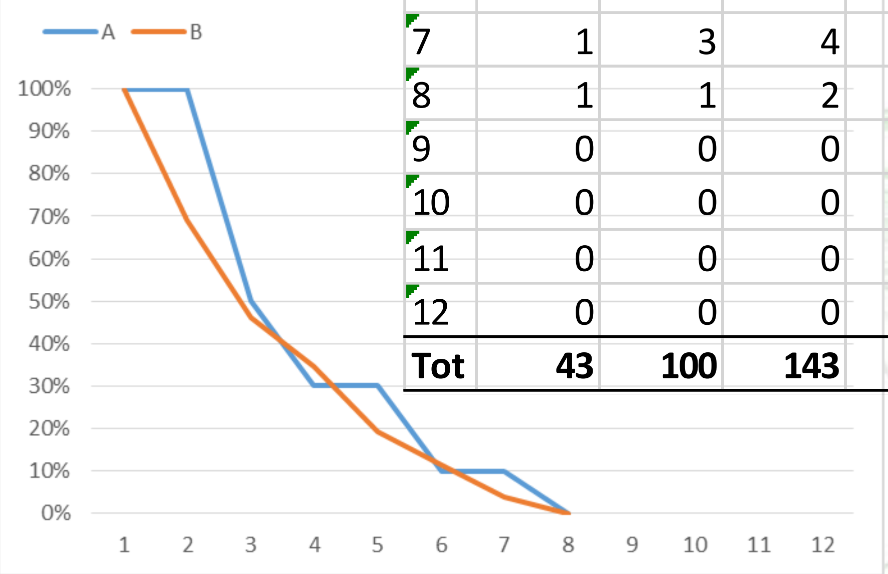
- Log Rank: calculate χ^2 with $M-1$ degrees of freedom for M subpopulations
- H_0: subpopulations $i=1,M$ have the same survival from the total population $All = \sum_i$
- $\chi^2 = \sum_i \frac{(Obs_i - Exp_i)^2}{Exp_i}$
$$= \sum_i \frac{\left(\sum_j D_{j,i} - \sum_j \left(E_{j,i} \cdot D_{j,All}/E_{j,All}\right)\right)^2}{\sum_j \left(E_{j,i} \cdot D_{j,All}/E_{j,All}\right)}$$
- Exp_i calculated at each time t_j
- $Obs_i = \sum_j D_{j,i} = D_i$ - decrements for subpopulation i
- For 1 DoF at 5% reject H_0 if $\chi^2 > 3.84$
- Only tests significance, does not quantify

Insight Decision Solutions Inc.

Single Factor Example 1

Pop	Mean	Rel
A	3.30	111%
B	2.85	96%
Tot	**2.97**	**100%**

t	Et A	Et B	Et	Dt A	Dt B	Dt	Exp A	Exp B
1	10	26	36	0	0	0	0.00	0.00
2	10	26	36	0	8	8	2.22	5.78
3	10	18	28	5	6	11	3.93	7.07
4	5	12	17	2	3	5	1.47	3.53
5	3	9	12	0	4	4	1.00	3.00
6	3	5	8	2	2	4	1.50	2.50
7	1	3	4	0	2	2	0.50	1.50
8	1	1	2	1	1	2	1.00	1.00
9	0	0	0	0	0	0		
10	0	0	0	0	0	0		
11	0	0	0	0	0	0		
12	0	0	0	0	0	0		
Tot	**43**	**100**	**143**	**10**	**26**	**36**	**11.62**	**24.38**

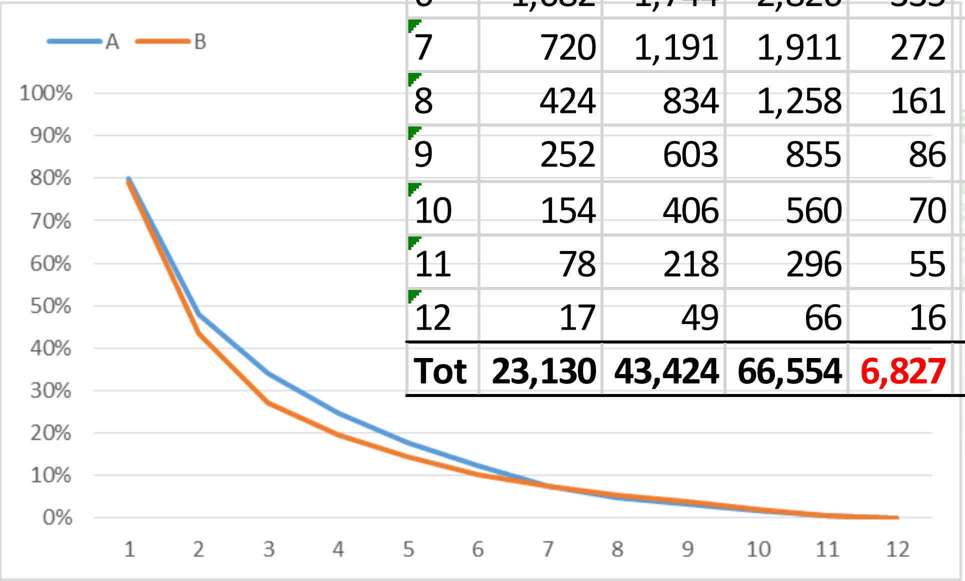


$$\chi^2 = \frac{(10 - 11.62)^2}{11.62} + \frac{(26 - 24.38)^2}{24.38} = 0.33$$

Single Factor Example 2

Pop	Mean	Rel
A	2.34	106%
B	2.13	97%
Tot	**2.20**	**100%**

t	Et A	Et B	Et	Dt A	Dt B	Dt	Exp A	Exp B
1	7,491	15,058	22,549	1,508	3,157	4,665	1,550	3,115
2	5,764	11,375	17,139	2,296	5,115	7,411	2,492	4,919
3	3,297	5,920	9,217	960	2,219	3,179	1,137	2,042
4	2,265	3,551	5,816	619	994	1,613	628	985
5	1,586	2,475	4,061	449	669	1,118	437	681
6	1,082	1,744	2,826	335	487	822	315	507
7	720	1,191	1,911	272	336	608	229	379
8	424	834	1,258	161	223	384	129	255
9	252	603	855	86	180	266	78	188
10	154	406	560	70	173	243	67	176
11	78	218	296	55	163	218	57	161
12	17	49	66	16	49	65	17	48
Tot	**23,130**	**43,424**	**66,554**	**6,827**	**13,765**	**20,592**	**7,137**	**13,455**



$$\chi^2 = \frac{(6{,}827 - 7{,}137)^2}{7{,}137} + \frac{(13{,}765 - 13{,}455)^2}{13{,}455} = 20.57$$

Hazard Ratio

- Hazard Ratio Estimate
  - $HR = \frac{Obs_A/Exp_A}{Obs_B/Exp_B}$
- Examples
  - Single Factor Ex 1: $\frac{10/11.62}{26/24.38} = 81\%$

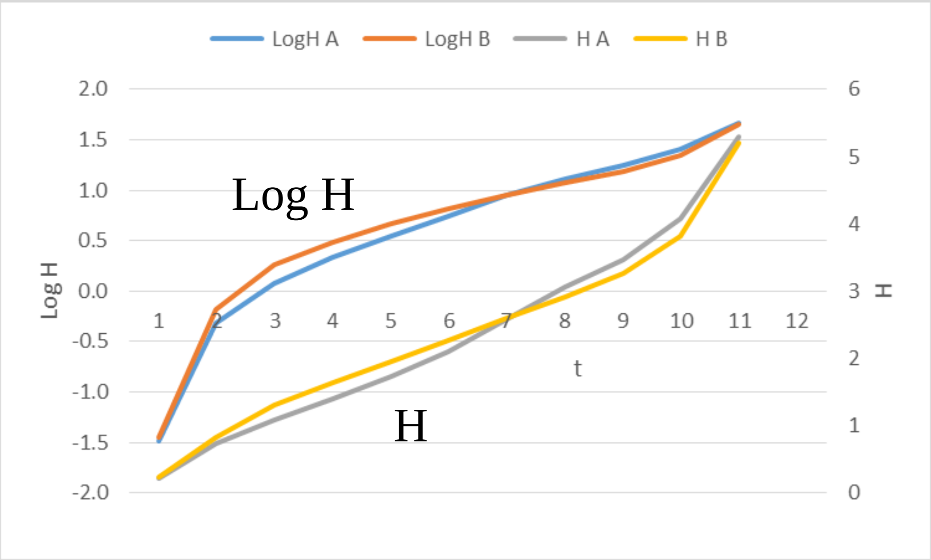
    Risk of closure is 19% lower for A compared to B but is not credible
  - Single Factor Ex 2: $\frac{6{,}827/7{,}137}{13{,}765/13{,}455} = 94\%$

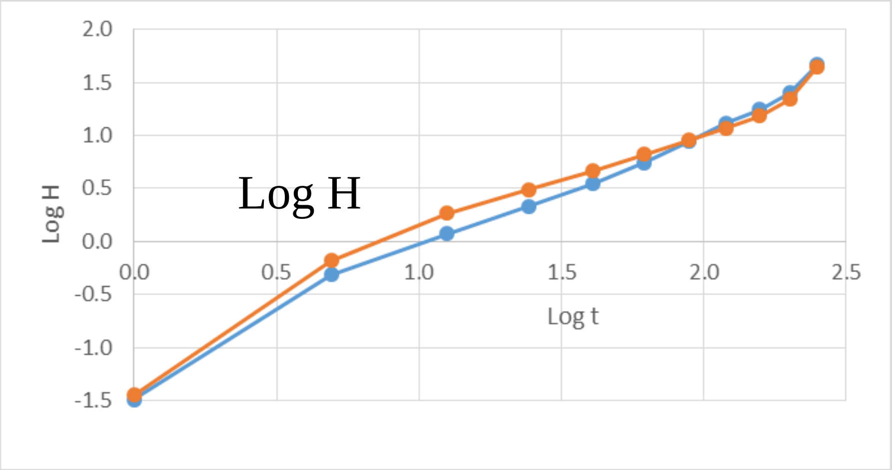
    Risk of closure is 6% lower for A compared to B and is credible

Cox Proportional Hazards

- Multiple risk factor test
- Quantifies survival behaviour
- Regression:
  - $h(t) = h_0(t)e^{(\sum_i \beta_i Z_i)}$
  - $\log(h(t)/h_0(t)) = \sum_i \beta_i Z_i$
- Assumes
  - Hazards are proportional
  - Hazard ratio is constant
- If not proportional for a variable, define as separate populations or strata and investigate separately for each

Testing Proportional Hazards





Plot $log\left(-log(S(t))\right)$ against $log(t)$

t	H A	H B	LogH A	LogH B
1	0.22	0.24	-1.49	-1.45
2	0.73	0.83	-0.31	-0.18
3	1.08	1.30	0.07	0.26
4	1.40	1.63	0.33	0.49
5	1.73	1.95	0.55	0.67
6	2.10	2.27	0.74	0.82
7	2.57	2.60	0.95	0.96
8	3.05	2.92	1.12	1.07
9	3.47	3.27	1.24	1.18
10	4.08	3.83	1.40	1.34
11	5.30	5.20	1.67	1.65
12				

Insight Decision Solutions Inc.

Closing

TRANSACTIONS OF SOCIETY OF ACTUARIES
1993 VOL. 45

MORTALITY DIFFERENCES BY HANDEDNESS: SURVIVAL ANALYSIS FOR A RIGHT-TRUNCATED SAMPLE OF BASEBALL PLAYERS

HARRY H. PANJER

- Reference
  - Chapter 19: Survival Models
  - Predictive Modeling Applications in Actuarial Science Vol I
  - Frees, Derrig, Meyers, 2014, Cambridge
- Contact
  - jmcgarry@insightdecision.com