Comments for MEDB 5502, Module 08, survival analysis

Topics to be covered

- What you will learn

  - A simple example of survival data

  - Overall Kaplan-Meier curve

  - The log rank test

  - The hazard function

  - The Cox regression model

  - Assumptions and data management

Survival analysis

- Time to event models

    - Death

    - Relapse

    - Rehospitalization

    - Failure of medical device

    - Pregnancy

- Not every patient experiences the event

    - These are censored observations

Survival analysis models are more properly called time to event models. You follow a group of patients from a certain time point and note the amount of time until they die.

Or the amount of time until they relapse. Or the amount of time until they need to be rehospitalized. Or the amount time until a medical device that you implanted in them fails.

I should note that while almost all of the events in a time to event model are bad, there are a few exceptions. In a study of couples with fertility problems, you might use a time to event model to study the time to pregnancy, a very happy event for any couple with fertility problems.

Mortality is the context under which time to event models were derived, so the term survival analysis has been used even when the event is different.

A key feature of survival analysis is that not every patient experiences the event. You should be glad that not everyone that you recruit for a clinical trial dies, but this adds a layer of complexity to the analysis.

The reasons for not experiencing the event can vary. Everyone dies, but maybe the event is death from cancer and if you patient gets hit by a bus, that patient does not experience the event. If the event is death from any cause, you still have to end the study in some time frame, and not everyone will die in that time frame. Not every patient gets rehospitalized, not every patient relapses, not every medical device fails.

A patient may drop out of a study, and you no longer are able to tell from that point onward whether that patient would have experienced the event sometime during the rest of your study.

When your patient does not die during the study, this is not a missing value. You have partial information. You know that the patient was alive for a certain amount of time. When you end the study within a certain time frame, you know that your patient dies at a time beyond the end date of your study. If your patient drops out after six months, you know that the patient survived for more than six months.

First fruit fly experiment, 1 of 4

```
data_dictionary: fly1.txt

description: |
  This dataset provides a simple example of what survival and censoring. It
provides an inuitive explanation of estimation of survival probabilities.

vars:
  day:
    label: Time until death
    unit: days
```

The following data represents survival time for a group of fruit flies and is a subset of a larger data set found at the Data and Story Library (DASL). The data set has been slightly modified to simplify some of these explanations.

There are 25 flies in the sample, with the first fly dying on day 37 and the last fly dying on day 96.

First fruit fly experiment, 2 of 4

37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, 70, 71, 72, 73, 75, 77, 79, 89, 94, 96

If you wanted to estimate the survival probability for this data, you would draw a curve that decreases by 4% (1/25) every time a fly dies.
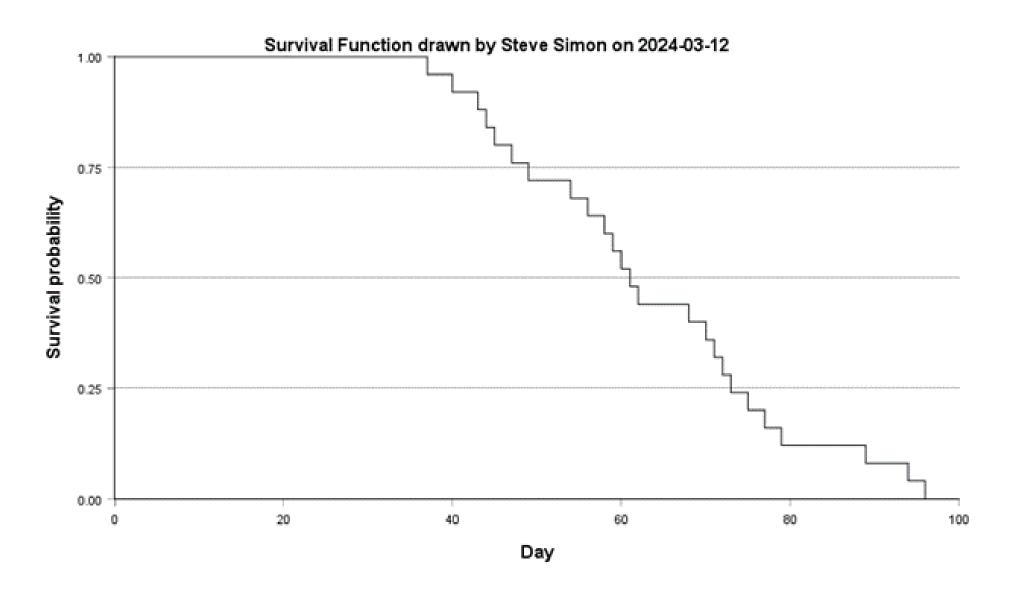
First fruit fly experiment, 3 of 4

	day	p			day	p			day	p
1	37	96%		10	58	60%		19	73	24%
2	40	92%		11	59	56%		20	75	20%
3	43	88%		12	60	52%		21	77	16%
4	44	84%		13	61	48%		22	79	12%
5	45	80%		14	62	44%		23	89	8%
6	47	76%		15	68	40%		24	94	4%
7	49	72%		16	70	36%		25	96	0%
8	54	68%		17	71	32%				
9	56	64%		18	72	28%				

The probability of survival drops by 4% (1/25) at each day of death.

First fruit fly experiment, 4 of 4

Survival Function drawn by Steve Simon on 2024-03-12

Here's a graph of these probabilities over time.

By tradition and for some rather technical reasons, you should use a stair step pattern rather than a diagonal line to connect adjacent survival probabilities.

Second fruit fly experiment, 1 of 4

37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, ??, ??, ??, ??, ??, ??, ??, ??, ??, ??

Now let's alter the experiment. Suppose that totally by accident, a technician leaves the screen cover open on day 70 and all the flies escape. This includes the fly who was going to die on the afternoon of the 70th day anyway. Oh the sadness of it all; the poor fly has the briefest of tastes of freedom then ends up shriveled up on a window sill.

You're probably worried that the whole experiment has been ruined. But don't be so pessimistic. You still have complete information on survival of the fruit flies up to their 70th day of life.

Second fruit fly experiment, 2 of 4

	day	event
1	37	1
2	40	1
3	43	1
4	44	1
5	45	1
6	47	1
7	49	1
8	54	1
9	56	1

	day	event
10	58	1
11	59	1
12	60	1
13	61	1
14	62	1
15	68	1
16	70	0
17	70	0
18	70	0

	day	event
19	70	0
20	70	0
21	70	0
22	70	0
23	70	0
24	70	0
25	70	0

Here's how you would code the data for importing into SPSS or any other software.
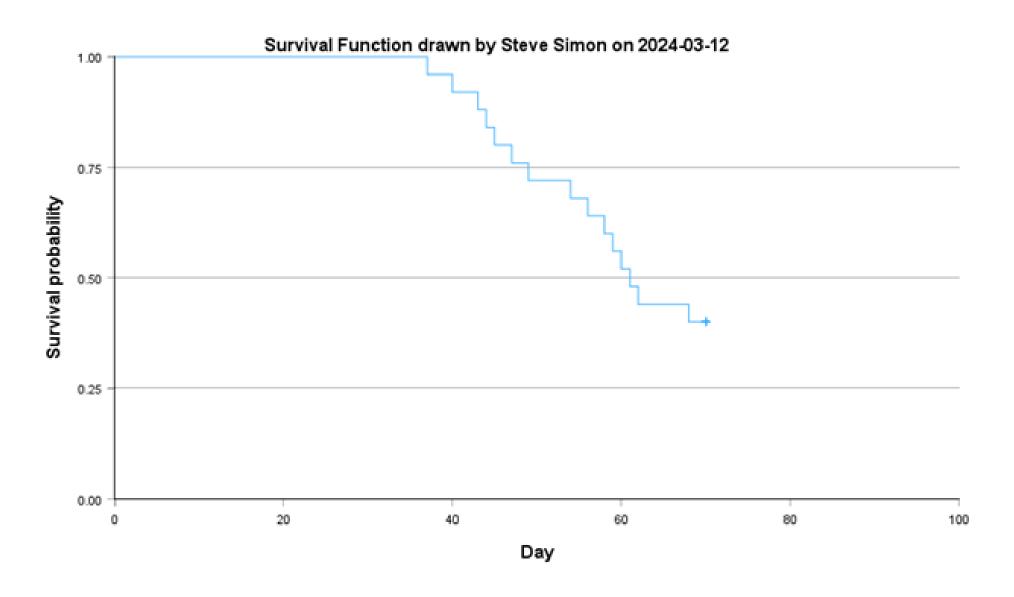
Second fruit fly experiment, 3 of 4

	day	event	p		day	event	p		day	event	p
1	37	1	96%	10	58	1	60%	19	70	0	
2	40	1	92%	11	59	1	56%	20	70	0	
3	43	1	88%	12	60	1	52%	21	70	0	
4	44	1	84%	13	61	1	48%	22	70	0	
5	45	1	80%	14	62	1	44%	23	70	0	
6	47	1	76%	15	68	1	40%	24	70	0	
7	49	1	72%	16	70	0		25	70	0	
8	54	1	68%	17	70	0					
9	56	1	64%	18	70	0					

We clearly have enough data to make several important statements about survival probability. For example, the median survival time is 61 days because roughly half of the flies had died before this day.

Second fruit fly experiment, 4 of 4

Here is a graph of the survival probabilities of the second experiment. This graph is identical to the graph in the first experiment up to day 70 after which you can no longer estimate survival probabilities.

By the way, you might be tempted to ignore the ten flies who escaped. But that would seriously bias your results. All of these flies were tough and hardy flies who lived well beyond the median day of death. If you pretended that they didn't exist, you would seriously underestimate the survival probabilities. The median survival time, for example, of the 15 flies who did not escape, for example, is only 54 days which is much smaller than the actual median.

Third fruit fly experiment, 1 of 4

37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, ??, 71, ??, ??, 75, ??, ??, 89, ??, 96

Let's look at a third experiment, where the screen cover is left open and all but four of the remaining flies escape. It turns out that those four remaining flies who didn't bug out will allow us to still get reasonable estimates of survival probabilities beyond 70 days.

Third fruit fly experiment, 2 of 4

	day	event
1	37	1
2	40	1
3	43	1
4	44	1
5	45	1
6	47	1
7	49	1
8	54	1
9	56	1

	day	event
10	58	1
11	59	1
12	60	1
13	61	1
14	62	1
15	68	1
16	70	0
17	71	1
18	70	0

	day	event
19	70	0
20	75	1
21	70	0
22	70	0
23	89	1
24	70	0
25	96	1

Here is how you would code the data for importing into SPSS.

Third fruit fly experiment, 3 of 4

	day	event	p			day	event	p			day	event	p
1	37	1	96%		10	58	1	60%		19	70	0	
2	40	1	92%		11	59	1	56%		20	75	1	20%
3	43	1	88%		12	60	1	52%		21	70	0	
4	44	1	84%		13	61	1	48%		22	70	0	
5	45	1	80%		14	62	1	44%		23	89	1	10%
6	47	1	76%		15	68	1	40%		24	70	0	
7	49	1	72%		16	70	0			25	96	1	0%
8	54	1	68%		17	71	1	30%					
9	56	1	64%		18	70	0						

What you do with the six escaped flies is to allocate their survival probabilities equally among the four flies who didn't bug out. This places a great responsibility among each of those four remaining flies since each one is now responsible for 10% of the remaining survival probability, their original 4% plus 6% more which represents a fourth of the 24% survival probability that was lost with the six escaping flies.

Another way of looking at this is that the six flies who escaped influence the denominator of the survival probabilities up to day 70 and then totally drop out of the calculations for any further survival probabilities. Because the denominator has been reduced, the jumps at each remaining death are much larger.

Third fruit fly experiment, 4 of 4

Survival Function drawn by Steve Simon on 2024-03-12

Here is a graph of the survival probability estimates from the third experiment.

These survival probabilities differ only slightly from the survival probabilities in the original experiment. This works out because the mechanism that caused us to lose information on six of the fruit flies was independent of their ultimate survival.

If the censoring mechanism were somehow related to survival prognosis, then you would have the possibility of serious bias in your estimates. Suppose for example, that only the toughest of flies (those with the most days left in their short lives) would have been able to escape. The flies destined to die on days 70, 71, 72, and 73, were already on their deathbeds and unable to fly at all, much les make a difficult escape. Then these censored values would not be randomly interspersed among the remaining survival times, but would constitute some of the larger values. But since these larger values would remain unobserved, you would underestimate survival probabilities beyond the 70th day.

This is known as informative censoring, and it happens more often that you might expect. Suppose someone drops out of a cancer mortality study because they are abandoning the drugs being studied in favor of laetrile treatments down in Mexico. Usually, this is a sign that the current drugs are not working well, so a censored observation here might represent a patient with a poorer prognosis. Excluding these patients would lead to an overestimate of survival probabilities.

Interpreting Kaplan-Meier plots

Survival Function drawn by Steve Simon on 2024-03-12
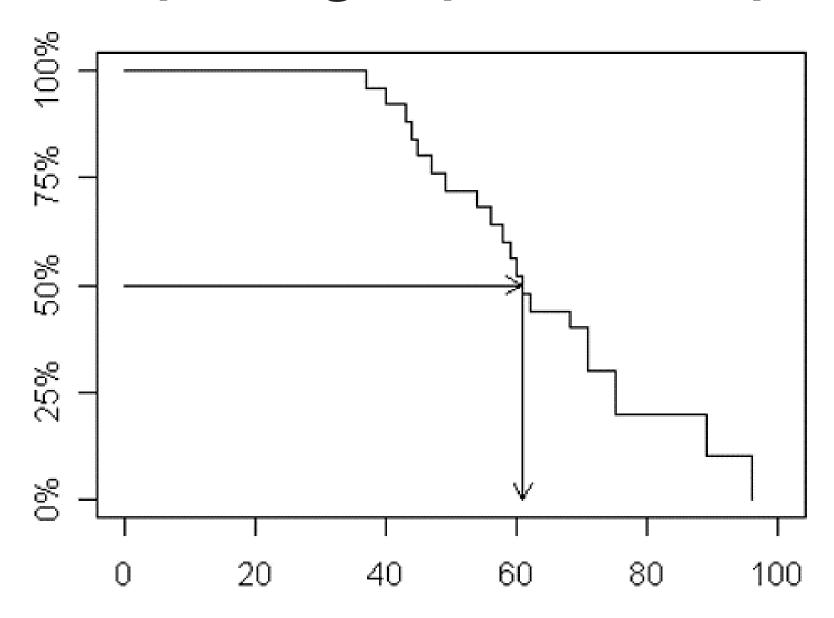
Happy corner

Sad corner

16

When you see a survival curve in a research paper, there are three ways to interpret it.

First, presuming that the event in question is a sad event (such as death, relapse), then the upper right hand corner is the happy corner. Most of you patients go for a very long time with only a small proportion suffering the negative event.
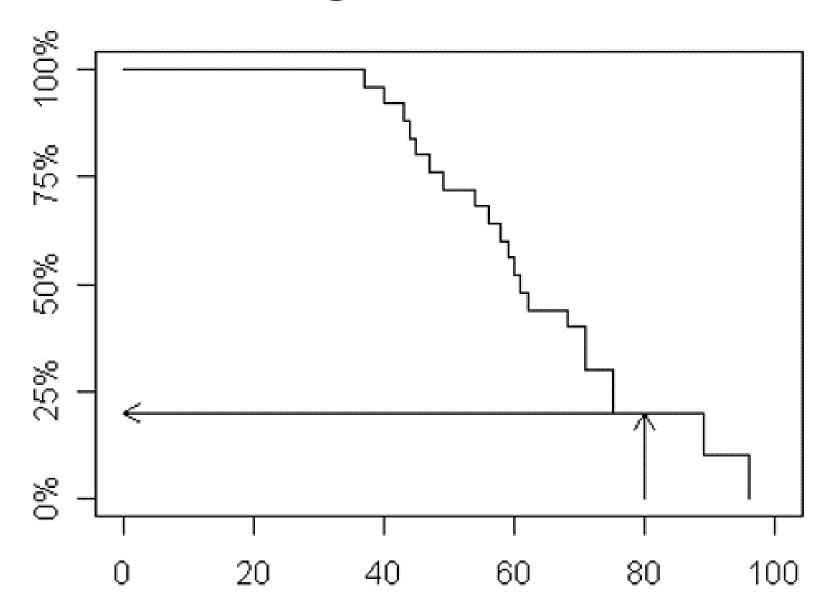
In contrast, the lower left corner is the sad corner. Most of your patients experience the bad event, and they experience it very quickly.

Interpreting Kaplan-Meier plots, 2 of 3

Next, you can get an estimate of the median (or other percentiles) by projecting horizontally until you intersect with the survival curve and then head down to get your estimate. In the survival curve we have just looked at, you would estimate the median survival as slightly more than 60 days.

You can also estimate probabilities for survival at any given time by projecting up from the time and then moving to the left to estimate the probability. In the example below, you can see that the 80 day survival probability is a little bit less than 25%.

Break #1

- What you have learned

  - A simple example of survival data

- What's coming next

  - Overall Kaplan-Meier curve

Worcester Heart Attack Study, 1 of 9

```
data_dictionary: whas500.dat

description: |
  The data represents survival times for a 500 patient  subset of data from
  the Worcester Heart Attack Study. You can find more information about this
  data set in Chapter 1 of Hosmer, Lemeshow, and May.
```

```
id:
  label: a sequential code from 1 to 100
age:
  label: Age at Admission
  unit: years
gender:
  value:
    Male: 0
    Female: 1
```

Worcester Heart Attack Study, 3 of 9

```
hr:
  label: Initial Heart Rate
  unit: Beats per minute
sysbp:
  label: Initial Systolic Blood Pressure
  unit: mmHg
diasbp:
  label: Initial Diastolic Blood Pressure
  unit: mmHg
```

```
bmi:
  label: Body Mass Index
  unit: kg/m^2
cvd:
  label: History of Cardiovascular Disease
  value:
    'FALSE': 0
    'TRUE': 1
afb:
  label: Atrial Fibrillation
  value:
    'FALSE': 0
    'TRUE': 1
```

```
sho:
  label: Cardiogenic Shock
  value:
    'FALSE': 0
    'TRUE': 1
chf:
  label: Congestive Heart Complications
  value:
    'FALSE': 0
    'TRUE': 1
av3:
  label: Complete Heart Block
  value:
    'FALSE': 0
    'TRUE': 1
```

```
miord:
  label: MI Order
  value:
    First: 0
    Recurrent: 1
mitype:
  label: MI Type
  value:
    non Q-wave: 0
    Q-wave: 1
```

```
year:
  label: Cohort Year
  value:
    yr1997: 1
    yr1999: 2
    yr2001: 3
admitdate:
  label: Admission Date
  format: mm/dd/yyyy
disdate:
  label: Hospital Discharge Date
  format: mm/dd/yyyy
```

```
fdate:
  label: Date of last Follow Up
los:
  label: Length of Hospital Stay
  unit: Days
dstat:
  label: Discharge Status from Hospital
  value:
    Alive: 0
    Dead: 1
```

```
lenfol:
  label: Follow Up Time
  unit: days
fstat:
  label: Vital Satus
  value:
    Alive: 0
    Dead: 1
```

Event count

fstat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	285	57.0	57.0	57.0
	1	215	43.0	43.0	100.0
	Total	500	100.0	100.0	

Overall Kaplan-Meier curve

Survival Function drawn by Steve Simon on 2024-03-10

Live demo, Overall Kaplan-Meier curve

Break #2

- What you have learned
  - Overall Kaplan-Meier curve
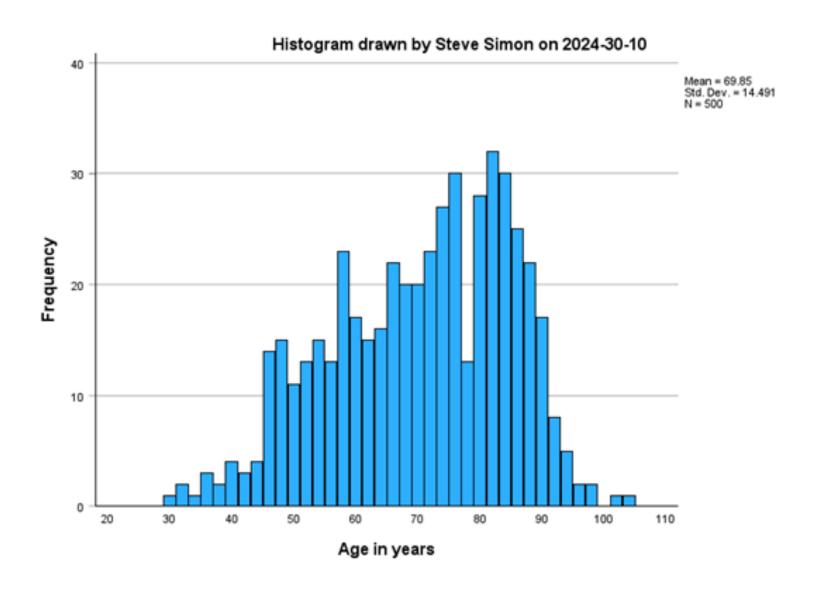- What's coming next
  - The log rank test

Event count by gender

gender * fstat Crosstabulation

Count

		fstat		
		0	1	Total
gender	0	189	111	300
	1	96	104	200
Total		285	215	500

Histogram of ages

Histogram drawn by Steve Simon on 2024-30-10

Mean = 69.85
Std. Dev. = 14.491
N = 500

Quality check of age group coding

age * age_group Crosstabulation

Count

age			age_group			Total
		1	2	3	4	
age	30	1	0	0	0	1
	32	2	0	0	0	2
	33	1	0	0	0	1
	35	1	0	0	0	1
	36	2	0	0	0	2
	37	2	0	0	0	2
	39	2	0	0	0	2
	40	2	0	0	0	2
	41	2	0	0	0	2
	42	1	0	0	0	1
	43	2	0	0	0	2
	44	2	0	0	0	2
	45	4	0	0	0	4
	46	10	0	0	0	10
	47	4	0	0	0	4
	48	11	0	0	0	11
	49	4	0	0	0	4
	50	7	0	0	0	7
	51	0	4	0	0	4
	52	0	9	0	0	9

35

Event count by age group, 1 of 2

age_group * fstat Crosstabulation

Count

		fstat		Total
		0	1	
age_group	1	55	5	60
	2	100	25	125
	3	92	78	170
	4	38	107	145
Total		285	215	500

Event count by age group, 2 of 2

age_group * fstat Crosstabulation

Count

		fstat		Total
		0	1	
age_group	65 and under	155	30	185
	66 to 80	92	78	170
	81 and above	38	107	145
Total		285	215	500

Kaplan-Meier analysis by gender, 1 of 3

Case Processing Summary

gender	Total N	N of Events	Censored N	Censored Percent
0	300	111	189	63.0%
1	200	104	96	48.0%
Overall	500	215	285	57.0%

Means and Medians for Survival Time

gender	Mean[a] Estimate	Std. Error	95% Confidence Interval Lower Bound	Upper Bound	Median Estimate	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
0	1448.506	55.852	1339.035	1557.976	2160.000	.	.	.
1	1260.208	75.267	1112.684	1407.732	1317.000	177.039	970.004	1663.996
Overall	1417.215	48.137	1322.867	1511.562	1627.000	159.555	1314.271	1939.729

a. Estimation is limited to the largest survival time if it is censored.

Kaplan-Meier analysis by gender, 2 of 3

Percentiles

gender	25.0% Estimate	25.0% Std. Error	50.0% Estimate	50.0% Std. Error	75.0% Estimate	75.0% Std. Error
0			2160.000	.	354.000	109.695
1	2353.000	176.168	1317.000	177.039	151.000	84.201
Overall	2353.000	79.465	1627.000	159.555	295.000	63.223

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	7.791	1	.005

Test of equality of survival distributions for the different levels of gender.
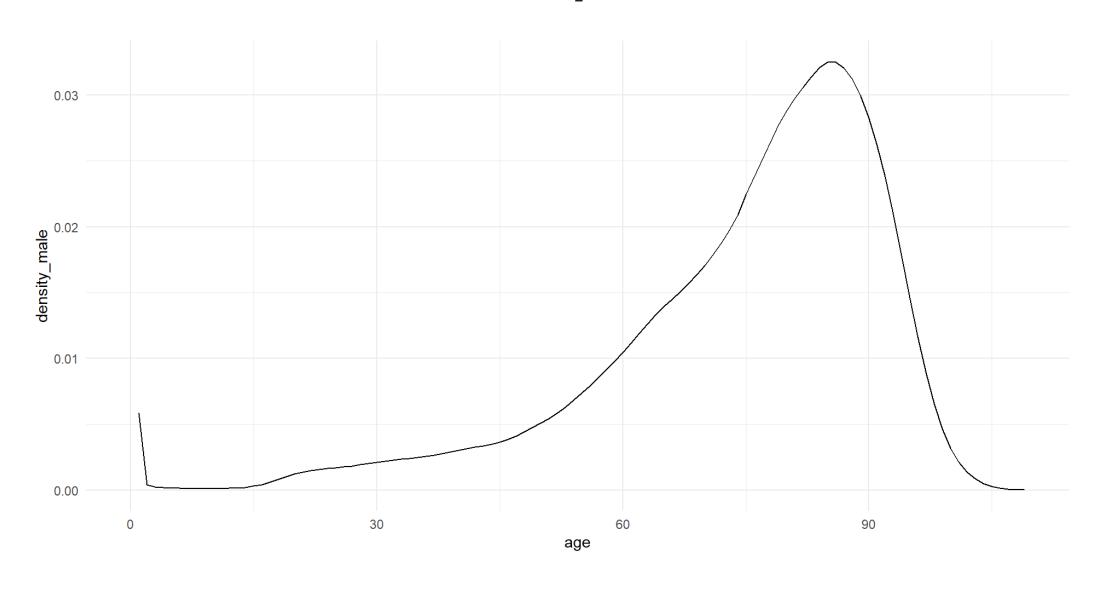
Kaplan-Meier analysis by gender, 3 of 3

Kaplan-Meier analysis by age group, 1 of 3

Case Processing Summary

age_group	Total N	N of Events	Censored N	Percent
65 and under	185	30	155	83.8%
66 to 80	170	78	92	54.1%
81 and above	145	107	38	26.2%
Overall	500	215	285	57.0%

Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
age_group	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
65 and under	2005.895	61.543	1885.270	2126.520	2353.000	550.241	1274.527	3431.473
66 to 80	1398.427	79.795	1242.028	1554.825	1624.000	259.869	1114.658	2133.342
81 and above	727.181	71.645	586.757	867.605	343.000	88.116	170.292	515.708
Overall	1417.215	48.137	1322.867	1511.562	1627.000	159.555	1314.271	1939.729

a. Estimation is limited to the largest survival time if it is censored.

Kaplan-Meier analysis by age group, 2 of 3

Percentiles

age_group	25.0%		50.0%		75.0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
65 and under	2358.000	.	2353.000	550.241	2353.000	574.473
66 to 80	2350.000	.	1624.000	259.869	345.000	126.261
81 and above	1359.000	181.588	343.000	88.116	33.000	12.970
Overall	2353.000	79.465	1627.000	159.555	295.000	63.223

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	129.254	2	<.001

Test of equality of survival distributions for the different levels of age_group.

Kaplan-Meier analysis by age group, 3 of 3

Live demo, The log rank test

Break #3

- What you have learned
  - The log rank test
- What's coming next
  - The hazard function

Life insurance example

I found some data on mortality from the Social Security website and plotted an approximation to the probability density function. There is an unusual early peak in this function because the first year of your life is one of the most dangerous ones you will have to face.
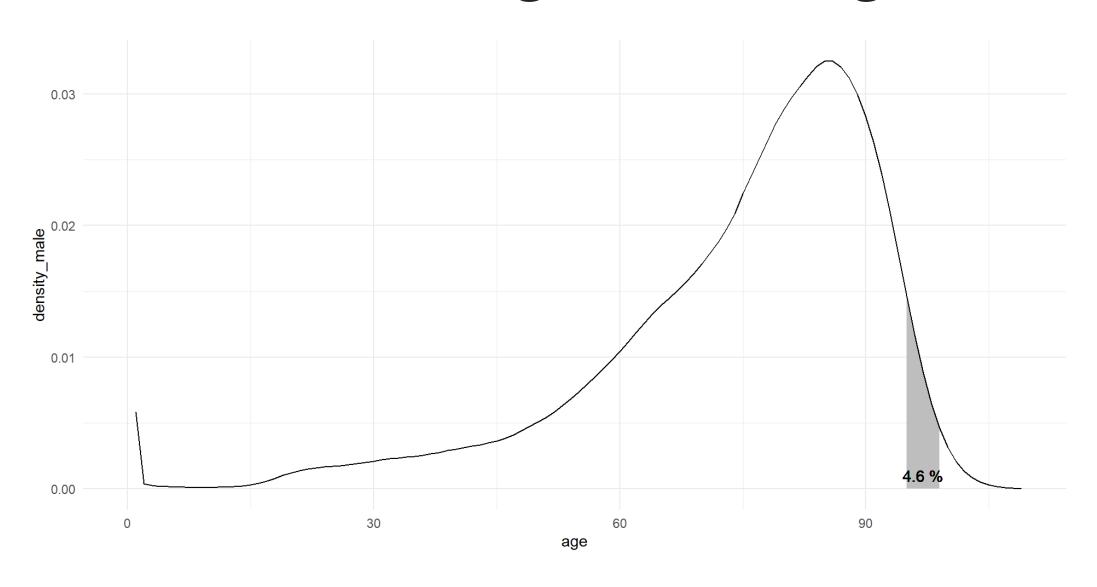
Imagine yourself working in life insurance sales. You want to price your policies so that you only ask for low payments on the policy when the risk of death is low. So let's calculate some probabilities.

Probabilities for ages 21 through 41

The probability of a potential customer dying between the ages of 21 and 41 is 0.04638.

Probabilities for ages 95 through 99

The probability of a potential customer dying between the ages of 95 and 99 is about the same, 0.04626. So should you charge the same amount for an insurance policy for someone 21 years old and someone 95 years old?

Why are these probabilities not comparable?

- Unequal time intervals

  - Fix by computing a rate

- Non-uniform probabilities over the interval

  - Fix by looking at narrow interval

- No adjustment for survivorship

  - Fix by dividing by survival probabilty

Obviously not. There are three things you need to fix first.

The most obvious flaw is the unequal time intervals, 20 years for the first probability and 4 years for the second probability.

You can fix this by computing a rate. You get the rate by dividing the probability by the width of the time interval.

The second flaw is that the probability changes over the interval, increasing in the first case and decreasing in the second case.

You can fix this by shrinking the width of the time interval.

The third flaw is a bit more subtle. The probability of dying between the ages of 95 and 99 are probabilities computed from the perspective of a newborn child. That probability is small not because the chances of dying are small at that age, but because so many have died before their 95th birthday.

If you are in insurance sales, you do not sell policies to newborn infants. You sell to people who have survived to a certain age. No one rises from their grave on their 95th birthday and asks for an insurance policy. First, because zombies aren't real, and second the zombie who died prior to year 95 would not be able to collect on an insurance policy that paid off for a death between 95 and 99.
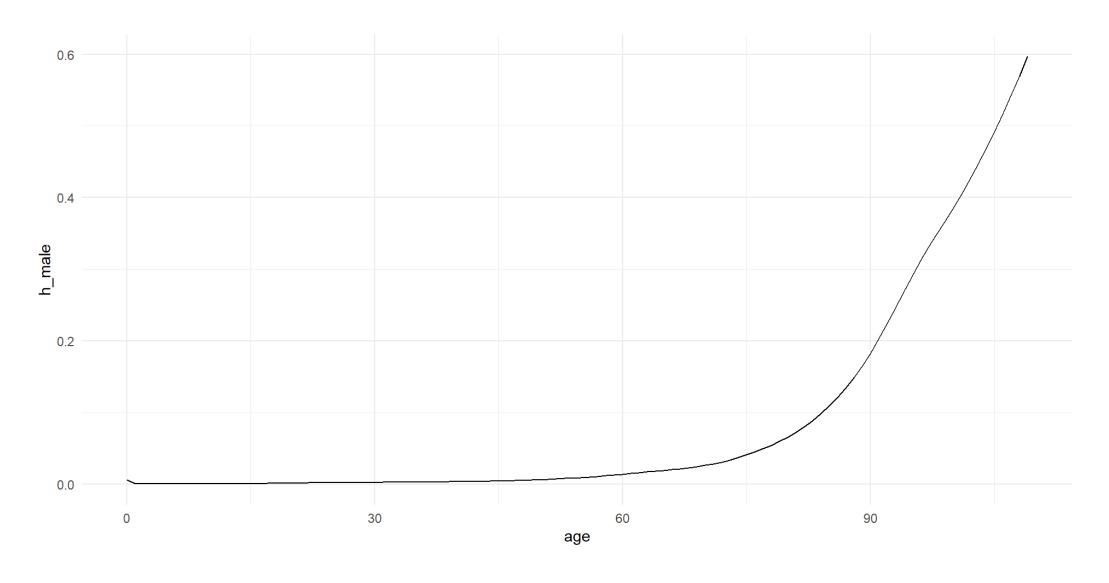
You can fix this by dividing by the survivor probability.

Hazard function, definition

-

- ■ where is the density function, and
  ■ is the survival function ()
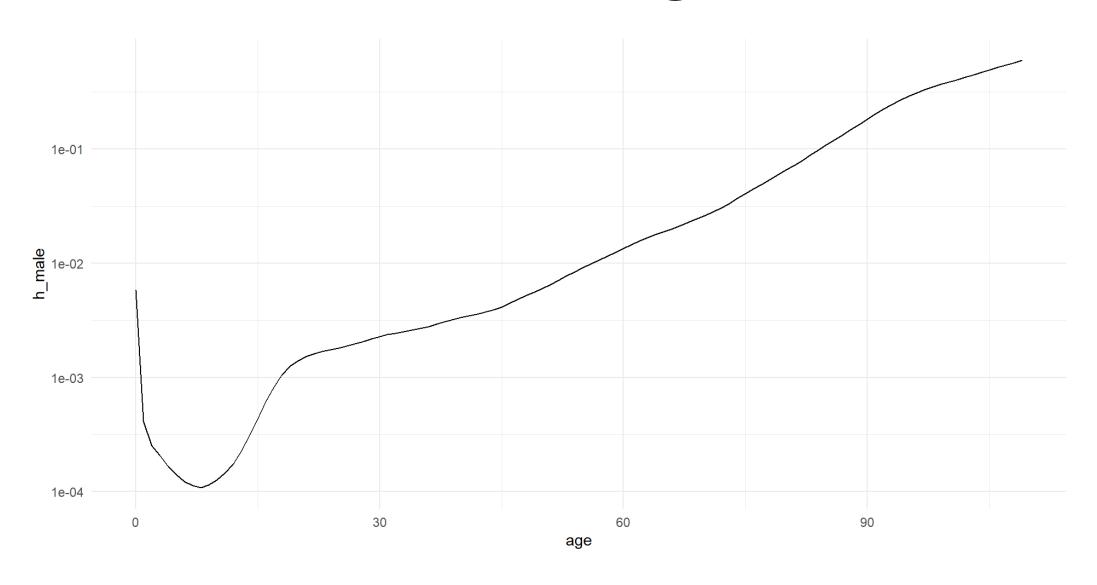
The hazard function addresses all three of the concerns mentioned above. It computes a rate by dividing by Δt. It shrinks the interval but using a limit. And it adjusts for survivorship by dividing by the survivor probability.

Hazard function, example

This is what the hazard function for mortality data looks like.

Hazard function on a log scale

The pattern becomes a bit clearer when you look at the hazard function on a log scale. The risk of death is high early in your life, but drops. There is a safe period during your pre-teen and early teen years, but then the risk rises because of an increase in deaths associated with things like driving, alcohol, and other drugs. Some of that fades as you mature but other risks increase because of the unavoidable aging of your body.

Break #4

- What you have learned

  - The hazard function

- What's coming next

  - The Cox regression model

Mean ages for men and women

Report

age

gender	Mean	N	Std. Deviation
Male	66.60	300	14.943
Female	74.72	200	12.301
Total	69.85	500	14.491

Unadjusted and adjusted Cox regression models for gender

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B) Lower	Upper
gender	.381	.138	7.679	1	.006	1.464	1.118	1.917

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B) Lower	Upper
gender	-.066	.141	.217	1	.641	.937	.711	1.234
age	.067	.006	116.401	1	<.001	1.069	1.056	1.082

Live demo, The Cox regression model

Break #5

- What you have learned

  - The Cox regression model

- What's coming next

  - Assumptions and data management

Assumptions of the log rank test

- Independence
  - From one patient to another
  - Of censoring mechanism

There is only one assumption in the log rank test, independence. There are, however, two dimensions of independence.

First, you have to assume that the chances of an event occurring for one patient is independent of the chances of that event for another patient. There is a famous example where this was not true. Carrie Fisher, an actor and author, died on December 27, 2016. Her mother, Debbie Reynolds, herself a famous actor, died on December 28, 2016. The two events were probably related, the stress of dealing with Carrie Fisher's death contributed to the death of Debbie Reynolds.

Often you can only assess the assumption of independence qualitatively, using your knowledge of the setting. Certain situations might make you concerned about the independence assumption, an infectious disease for example. If your data has clusters (family units, litter mates), then independence within a cluster might be questionable.

You also have to assume that when an event is censored, it is for a reason unrelated to what that event time might have been.
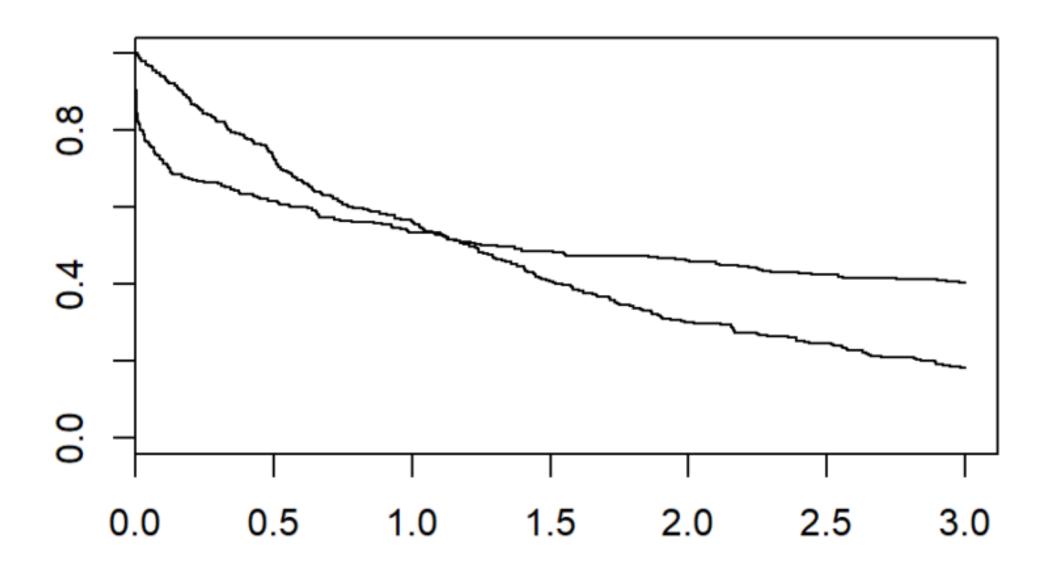
A patient may drop out of a study, for example, and you no longer know if or when the event occurred, except that it had to occur sometime after the date they dropped out. If a cancer patient drops out of a study because they went down to Mexico for an laetrile treatments, that might cause you to question the independence assumption. No one makes the trip to Mexico for an unapproved and unproven treatment if their current therapy was going well. The act of dropping out is informative as it indicates, at least in some cases, that death, relapse, or any other bad event is likely to happen sooner rather than later.
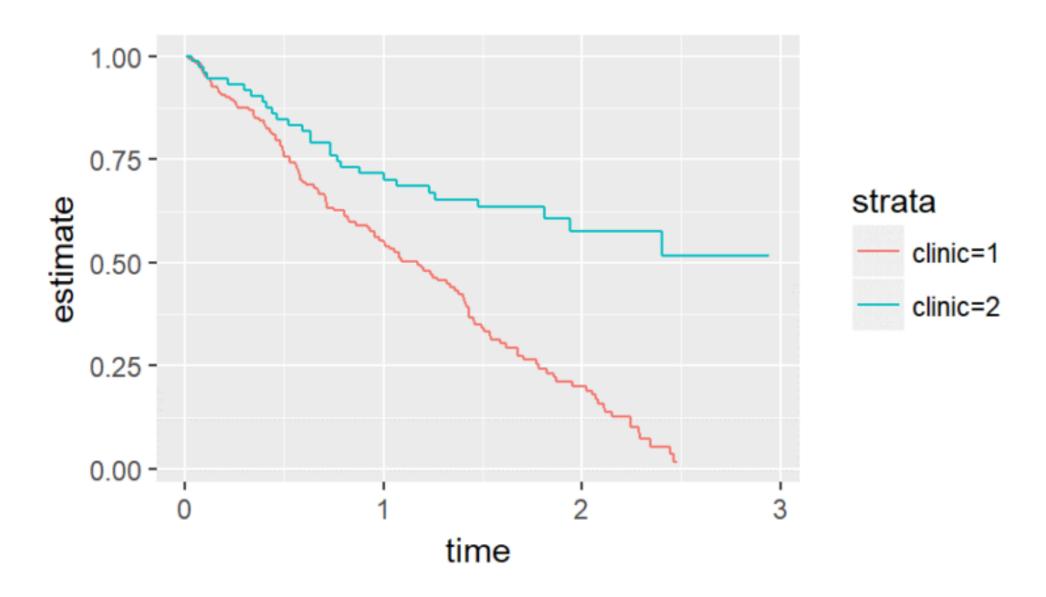
Assumptions of the Cox regression model

- Independence

- Proportional hazards assumption

- Possible violations of proportional hazards

  - Survival curves that cross

  - One curve flattening out over time
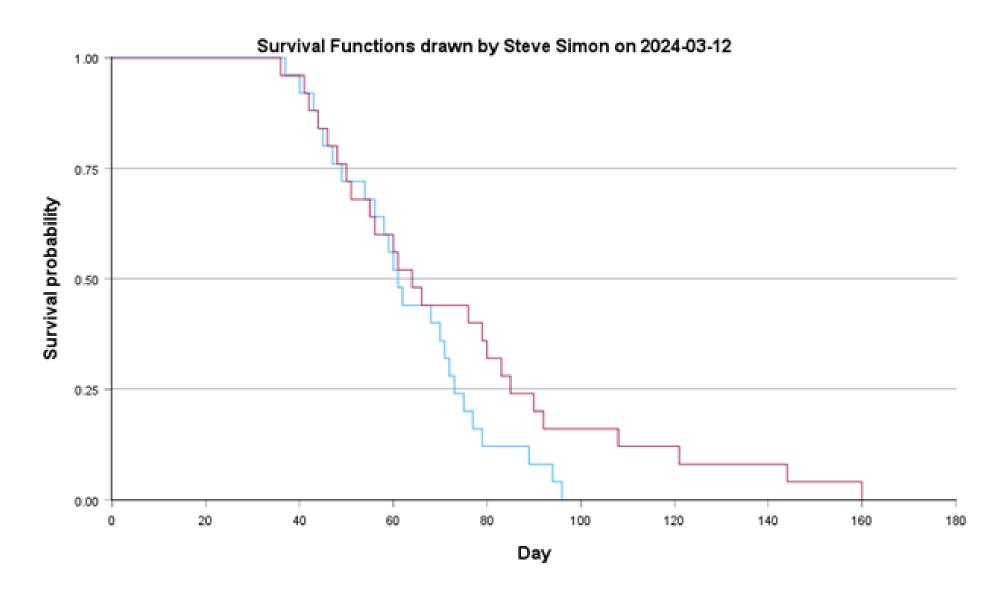
  - Curves diverge only at later times

Survival curves that cross

One curve flattening out over time

Curves diverge only at later times

Survival Functions drawn by Steve Simon on 2024-03-12

Sample size issues

- Rule of 50

- Rule of 15

Use ISO format for dates

Understand the internal storage system for dates

Date management

- The three dates you need

    - the date of origin,

    - the date of the event (if it occurred),

    - the date of last contact with the patient.

The date of origin

- Rehospitalization
  - use date of first discharge.
- Failure of a mechanical device
  - use date of implant.
- Divorce
  - use date of marriage.
- Loan default
  - use date of loan contract.
- Infectious disease
  - use date of first exposure.

You have various choices for the date of origin. It depends a lot on the context of the research.

The data of the event

- Define your event precisely

  - All-cause mortality

  - Mortality related to the health condition

  - Composite endpoints (e.g., death or relapse)

    - Requires comparing the earlier of two dates.

- If the event did NOT happen, leave this field blank/missing.

The date of last contact

- If event did not occur
    - Must be specified
    - Typically last medical exam or last telephone contact
- If event did occur
    - Make same as event date, or
    - Leave blank

Every patient will have a date of last contact. It will be the last time that you have been able to contact the patient and confirm that the event has not yet occurred.

If the event has occurred, you have several reasonable choices: put the event date in this field also, or leave the field blank/missing.

A date after the event date may represent a data error!

Survival calculations, 1 of 2

- Time = max(Date of event, Date of last contact) - Date of origin

- Censoring variable = 0 if Date of event is missing, 1 if not

Survival calculations, 2 of 2

patient	origin	event	last_contact	time	censored
1	2024-01-02		2024-01-30	28	0
2	2024-01-04	2024-01-12		8	1
3	2024-01-06		2024-01-30	24	0
4	2024-01-08		2024-01-30	22	0
5	2024-01-10		2024-01-30	20	0
6	2024-01-12		2024-01-30	18	0
7	2024-01-14	2024-01-15		1	1
8	2024-01-16	2024-01-28		12	1
9	2024-01-18	2024-01-25		7	1

Summary

- What you have learned
  - A simple example of survival data
  - Overall Kaplan-Meier curve
  - The log rank test
  - The hazard function
  - The Cox regression model
  - Assumptions and data management

Additional topics??

- Kaplan-Meyer - Survival Analysis
  - Define Survival Analysis
  - List the scientific applications of Survival Analysis
  - Know the general overview of Survival Analysis
  - Identify the type of data used in Survival Analysis
  - Identify data requirements for Survival Analysis
  - Identify the research design for a Survival Analysis
  - List the benefits of Survival Analysis
  - Define "censoring"
  - Define "event observation"
  - Outline the history of Survival Analysis
  - Define "Life Tables"
  - List the characteristics of Kaplan-Meier Survival Analysis
  - List and define the assumptions of The Kaplan-Meier method
  - Explain how to test Assumption #6 - Censorship equality
  - Define "left censoring"
  - Define "right censoring"
  - Define the 2 goals of the Kaplan-Meier method
  - Define the characteristic of the Event Status variable
  - Define the characteristic of the Time variable
  - Define the Subject Identifier variable
  - Define the Between Subjects Factor
  - Run a Kaplan-Meier analysis using SPSS
  - Define and explain the Log-Rank Test, the Breslow test and the Tarone-Ware test.
  - Interpret a Survival Functions Plot
  - Identify the median survival values on a Survival Functions Plot
  - Interpret the median values of a Kaplan-Meier Survival analysis
  - Identify the useful functions of a survival distribution
  - Explain the pair wise comparisons generated in a Kaplan-Meier survival analysis
  - Create a complete write-up of a Kaplan-Meier survival analysis
- Cox Regression

- Cox Regression
  - Define Cox Regression
  - Explain what a Cox Regression analysis is used for
  - Compare and contrast a Cox Regression analysis and a Kaplan-Meier survival analysis
  - Identify the benefits of a Cox Regression analysis
  - Define Hazard Function
  - Explain the Cox Regression Equation
  - Define "maximum Likelihood"
  - Interpret a positive regression coefficient and a negative regression coefficient
  - Compare and contrast the interpretation of a Cox Regression analysis and a Kaplan-Meier survival analysis
  - Identify the assumptions of a Cox Regression analysis
  - Define the Proportional Hazards assumption
  - Explain how to test the Proportional Hazards assumption
  - Explain what will occur if the Proportional Hazards assumption is violated
  - Run a Cox Regression analysis in SPSS
  - Interpret the Omnibus Tests of Model Coefficients
  - Interpret the Beta coefficients of a Cox Regression analysis
  - Interpret the Hazard Function Curve
  - Create a complete write-up of a Cox Regression analysis
- Learning objectives
  - Define Propensity Score
  - Define propensity Score Matching
  - Explain the main reason for the use of propensity scores
  - Explain the use of Propensity Scores for a Treatment-Control Study
  - List the 3 relationship between the covariate, the Outcome Variable and the Treatment assignment
  - Define the 3 relationship between the covariate, the Outcome Variable and the Treatment assignment
  - State the "biggest" challenge of Propensity Score calculation
  - Define the most important step in Propensity Score calculation
  - Explain the mathematical theory of the Propensity Score
  - List and define the Assumptions for Propensity Score calculation
  - Explain the problem of a Propensity Score of 0 or 1

- Explain the Area of Common Support
- Explain the need for an ideal number of covariates for the Propensity Score calculation
- Explain the selection of covariates for Propensity Score Calculation
- List and explain the methods for matching groups using Propensity Scores
- List and explain the limitations/concerns of using Propensity Scores
- Explain the benefits of Propensity Score Matching over using a regression model for covariate control.
- Explain the benefits of using a regression model over Propensity Score Matching for covariate control.