An Example of Survival Analysis Data Applied to Covid-19

Jorge Luis Romeu, Ph.D.
https://www.researchgate.net/profile/Jorge_Romeu
http://web.cortland.edu/romeu/
Email: romeu@cortland.edu

1.0 Introduction

The present *survival analysis* is intended to illustrate the power that this statistical technique can offer to Covid-19 data studies. It is part of *our pro-bono collaboration to the American struggle against Covid-19*, whereby retired professionals like myself would provide input, based on their long experience. You can find our *Proposal for Fighting Covid-19 and its Economic Fallout* in https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout

We have also written *Multivariate Statistics in the Analysis of Covid-19 Data* and its sequence, *More on Applying Multivariate Statistics to Covid-19 Data*, both of which can be found in: https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19 and https://www.researchgate.net/publication/342154667_More_on_Applying_Principal_Components_Discrimination_Analysis_to_Covid-19 These statistical methods provide useful tools for classification of states, regions, counties, etc. according to levels of infection and other metrics.

In addition, we have written a tutorial on the use of *Design of Experiments (DOE) Applied to the Assessment Covid-19*. It provides an example of a tool for assessing and controlling appropriate levels of infection in states and regions. It can also be found in our ResearchGate web Page: https://www.researchgate.net/publication/341532612_Example_of_a_DOE_Application_to_Coronavarius_Data_Analysis We have written an assessment of the results of 25 years off-shoring tens of thousands American jobs, and the impact this has had on US preparedness to fight the Coronavarus Pandemic, found in: https://www.researchgate.net/publication/341685776_Off-Shoring_Taxpayers_and_the_Coronavarus_Pandemic And we have written a short study on the use of reliability methods in the design and operation of ICU units, that can be found in: https://www.researchgate.net/publication/342449617_Example_of_the_Design_and_Operation_of_an_ICU_using_Reliability_Principles

In this paper we analyze a data set from patients in an ICU Ventilator[1]. We start our analysis by considering only the *Time to Death or Exit* (recovery) of a Ventilator patient. We then include covariates (Comorbidity information) for each Ventilator patient. We considered Patient age, the number of Comorbidities[2], and weather the end of their *Time in Ventilator* was caused by death, or by a recovery.

[1] In the next section we describe how such ICU Data Set was created, to help illustrate our survival analysis.
[2] Such as High Blood Pressure, Cholesterol, Diabetes, Heart Disease, etc.

Then, using such information, *we analyze* (1) **the probability of survival of a Patient, given the Time in a Ventilator, their age, and Comorbidities; and (2) the strength and direction (if any) of the effects of the different covariates, in the patient's ICU Ventilator sojourn**.

2.0 The Data

We have tried unsuccessfully to obtain ICU Ventilator patient data from several organizations, including the American Statistical Association (ASA) and NY Upstate Medical Center. But it has not been possible to acquire these. Since we believe it is important to illustrate the use of survival analysis techniques that can be implemented, and the power of their results, we have created one.

We have used a real data set taken from an organization that we belong to[3]. We have collected, starting on the year 2000, (1) the time some members spent in it, before leaving the organization, and (2) the time other members have spent, since they joined this organization, to date. As we do not know, a priori, whether members will or will not leave the organization in the future, the time as members of those still remaining in the organization can be considered *Censored* information.

We have created, using judgment and experience, the *number of comorbidities and age*, for each Ventilator patient. We aim to show what can be achieved using such *patient concomitant data*.

Such data gives this analysis a real flavor. *Time to leave the organization* may be considered as *ICU Ventilator Time to Death* of a patient. This data may also be considered in two ways: (1) as censored data, it is the time on Ventilators, up to the end of the present data collection (i.e. these *patients are still alive* and connected to the device). *Alternatively*, it can be considered (2) as *time on ICU Ventilators by patients that recovered*. In this *second interpretation, data comes from* a *different population* (of recovery). In either case, it is of interest to implement a survival analysis.

Table 1: Survival Analysis Data

Row	Time	Censor	Age	Comorb	LogTime
1	9	0	45	1	2.19722
2	13	0	50	1	2.56495
3	7	0	40	0	1.94591
4	7	0	43	0	1.94591
5	5	0	35	0	1.60944
6	6	0	38	0	1.79176
7	2	0	44	0	0.69315
8	15	0	47	1	2.70805
9	1	0	33	0	0.00000
10	10	0	48	1	2.30259
11	5	0	41	0	1.60944
12	14	0	49	1	2.63906
13	9	0	42	0	2.19722
14	1	0	45	0	0.00000
15	9	0	36	0	2.19722
16	9	0	39	0	2.19722
17	13	0	46	1	2.56495

[3] Analysis of Years to Demit. https://web.cortland.edu/matresearch/LodgAnalYrstoDemit.pdf

18	1	0	32	0	0.00000
19	2	0	41	0	0.69315
20	9	0	44	1	2.19722
21	8	0	42	1	2.07944
22	4	0	39	0	1.38629
23	9	0	51	1	2.19722
24	9	0	49	0	2.19722
25	6	0	42	0	1.79176
26	14	0	55	1	2.63906
27	3	0	30	0	1.09861
28	5	0	36	1	1.60944
29	5	0	38	1	1.60944
30	4	0	34	0	1.38629
31	6	0	39	1	1.79176
32	4	0	36	0	1.38629
33	9	0	45	1	2.19722
34	13	0	47	1	2.56495
35	7	0	42	0	1.94591
36	7	0	44	1	1.94591
37	3	0	36	0	1.09861
38	10	0	47	1	2.30259
39	8	0	41	0	2.07944
40	2	1	73	1	0.69315
41	2	1	58	2	0.69315
42	6	1	60	1	1.79176
43	5	1	52	0	1.60944
44	15	1	65	2	2.70805
45	1	1	72	1	0.00000
46	4	1	66	1	1.38629
47	8	1	61	1	2.07944
48	1	1	55	2	0.00000
49	1	1	63	2	0.00000
50	1	1	78	1	0.00000
51	5	1	73	2	1.60944
52	7	1	77	1	1.94591
53	4	1	79	1	1.38629
54	2	1	82	1	0.69315
55	5	1	73	1	1.60944
56	2	1	78	2	0.69315
57	5	1	74	1	1.60944
58	7	1	68	1	1.94591
59	13	1	66	1	2.56495
60	3	1	69	2	1.09861
61	8	1	77	0	2.07944
62	1	1	85	2	0.00000
63	11	1	55	1	2.39790
64	7	1	45	2	1.94591
65	5	1	49	2	1.60944
66	12	1	57	1	2.48491
67	9	1	51	1	2.19722
68	6	1	66	2	1.79176
69	4	1	69	2	1.38629

70	7	1	59	1	1.94591
71	9	1	55	2	2.19722
72	3	1	67	2	1.09861
73	11	1	59	1	2.39790
74	5	1	68	2	1.60944
75	4	1	72	1	1.38629
76	7	1	77	1	1.94591
77	6	1	73	1	1.79176
78	9	1	70	0	2.19722
79	4	1	79	1	1.38629
80	3	1	80	2	1.09861
81	3	1	82	2	1.09861
82	7	1	81	0	1.94591
83	4	1	84	1	1.38629
84	4	1	85	2	1.38629
85	7	1	72	1	1.94591
86	3	1	66	2	1.09861
87	5	1	69	2	1.60944
88	3	1	77	2	1.09861
89	5	1	79	0	1.60944
90	3	1	84	0	1.09861

End of Table #1

We consider *two hypothetical cases* in our analysis, regarding the interpretation of this data:

In the first case, we consider ICU Ventilator Failures as Deaths. The Censored data (*Code 0*) are considered as those *patients still on an ICU* Ventilator, *at the* time of the *end of this study*. We will *refer to it* in what follows as: *First Case*.

In the second case, we consider ICU Ventilator Failures also as Deaths. But we consider *Code 0* data as coming from a *different population*: that of *patients* that have *successfully recovered* and have left the ICU Ventilator. We will *refer to it* in what follows as: *Second Case*.

Analysis *Results* will be *different, but equally useful and applicable* to Covid-19 situations.

Finally, *Reliability is defined*, in general, as the Probability that an activity developed in time X exceeds a specific, pre-specified time T. In mathematical terms: $R(T) = P\{X>T\}$

In this paper, the activity is the Patient's time X in an ICU Ventilator; and the time T, is the time to *Failure* (Death or recovery of the Patient). Thence, we study the Reliability or Probability that a Patient time in an ICU Ventilator (sojourn), either by Death or recovery, exceeds time T.

3.0 Non Parametric Survival Analysis: First Case

We start by *considering the First Case*. Assume *we conduct an ICU study during twenty days, and compute Time in Ventilators for its 90 patients*. At the time the study is terminated, *39 are still alive -all other* patients having *passed away*. We *consider the 39 still alive as Censored*, as we do not know when they would have died, had our study continued beyond its 20 days.

We use the *Kaplan-Meier* Distribution Free survival procedure, as we *do not want to commit to any specific statistical distribution*. Analysis Results are given below:

Variable: Time in Ventilator

Censoring Information	Count
Uncensored value	51
Right censored value	39

Censoring value: Code = 0

Nonparametric Estimates:

Characteristics of the Variable

	Standard	95.0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
8.41259	**0.575924**	**7.28380**	**9.54138**

(Average sojourn, or Time in Ventilator)

Median = 7; IQR = 9; Q1 = 4; Q3 = 13 (Non Parametric Four Number Descriptors)

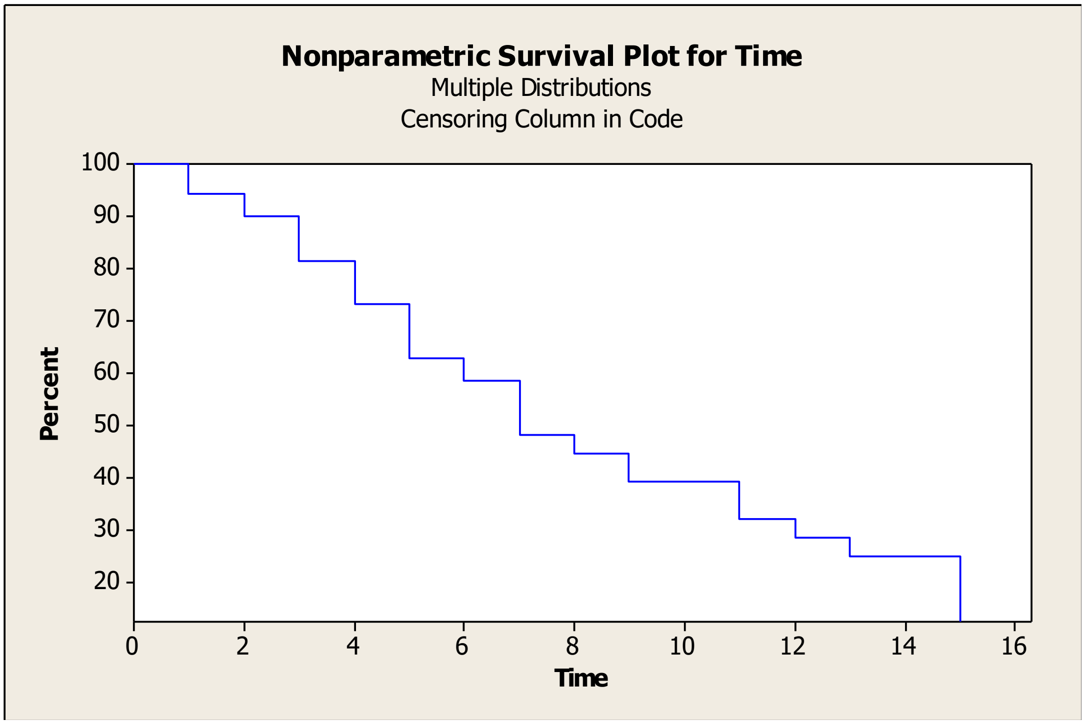
Table 2: Kaplan-Meier Estimates

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	Upper
1	90	5	0.944444	0.0241452	0.897121	0.991768
2	82	4	0.898374	0.0321285	0.835403	0.961345
3	76	7	0.815629	0.0416997	0.733899	0.897359
4	67	7	0.730414	0.0482026	0.635939	0.824889
5	57	8	0.627900	0.0533509	0.523334	0.732466
6	45	3	0.586040	0.0549965	0.478249	0.693831
7	39	7	0.480853	0.0577339	0.367697	0.594009
8	28	2	0.446507	0.0584957	0.331857	0.561156
9	24	3	0.390693	0.0594000	0.274271	0.507115
11	11	2	0.319658	0.0665299	0.189262	0.450054
12	9	1	0.284141	0.0679602	0.150941	0.417340
13	8	1	0.248623	0.0681170	0.115116	0.382130
15	2	1	0.124311	0.0942690	0.000000	0.309075

An example of interpretation of the above Reliability/Survivability table is as follows:

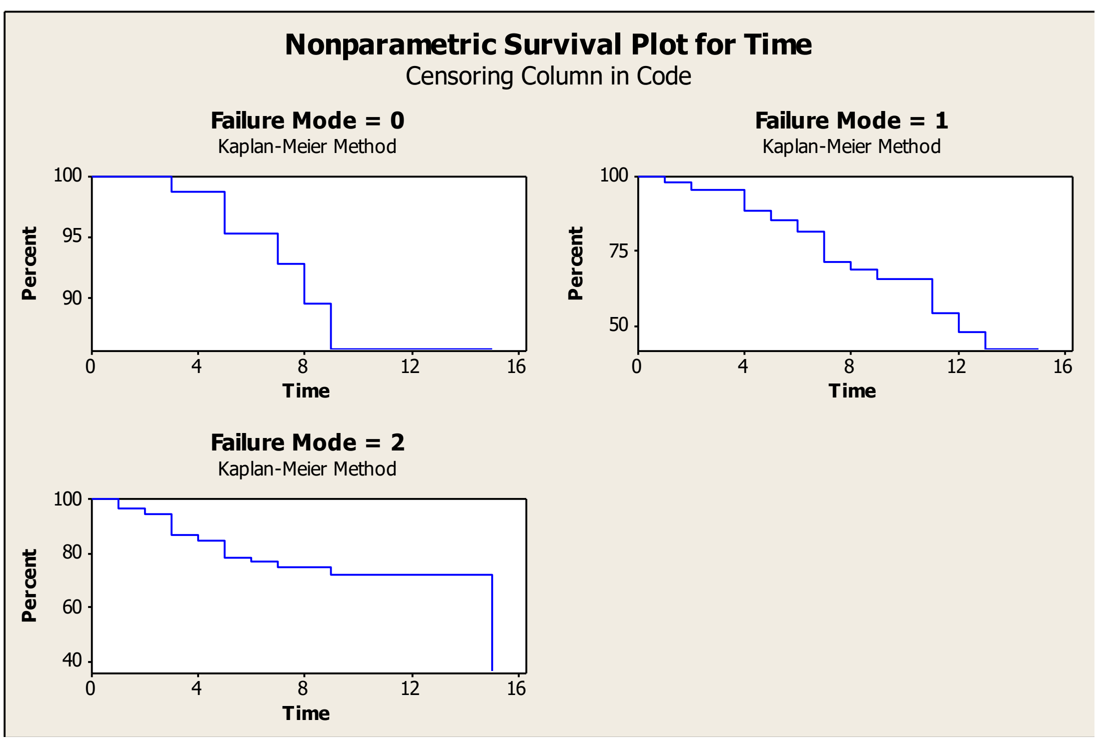
At the end of the first day, when the ICU Ventilator unit started with 90 patients, five had *failed*. The *Probability of Surviving One Day* is 0.94 or 94% (of failing: 1-0.94=0.06 or 6%). A *95% CI* (confidence interval) for this surviving probability is: 89.7% to 99.2%. There were 90 – 5 = 85 Patients surviving, but *only 82 were at risk at the end of Day 2.* This is *because there are three Censored Patients*. We can do likewise with all other days in the Table. For these results we assumed *Censoring*, and no effect of *Comorbidities*.

The graphs of these probabilities (survival plots) are shown next. These survival plots present, in addition, survival analyses that take into consideration the *Number of Comorbidities* of all these Patients (0, 1, 2). We also present the graphs of the corresponding Failure/Hazard Rates.



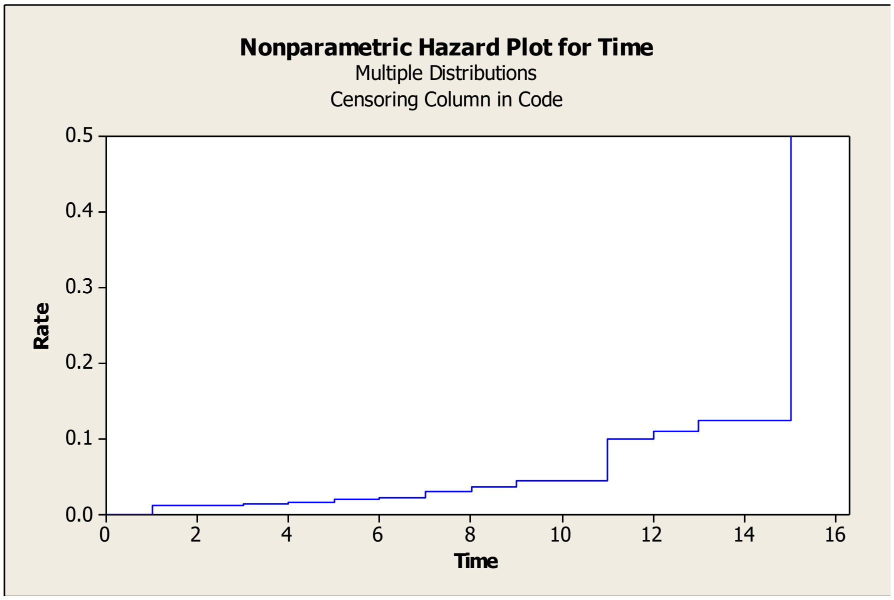
The above survival plot may be used to assess Patients just placed in an ICU Ventilator. For it provides an estimate of the probability of survival for any sojourn. For example, the probability of surviving six days in a Ventilator is 60%, and the probability of surviving twelve days is 30%.

Survival Plots, considering all three Comorbidities (Failure Modes 0,1,2)



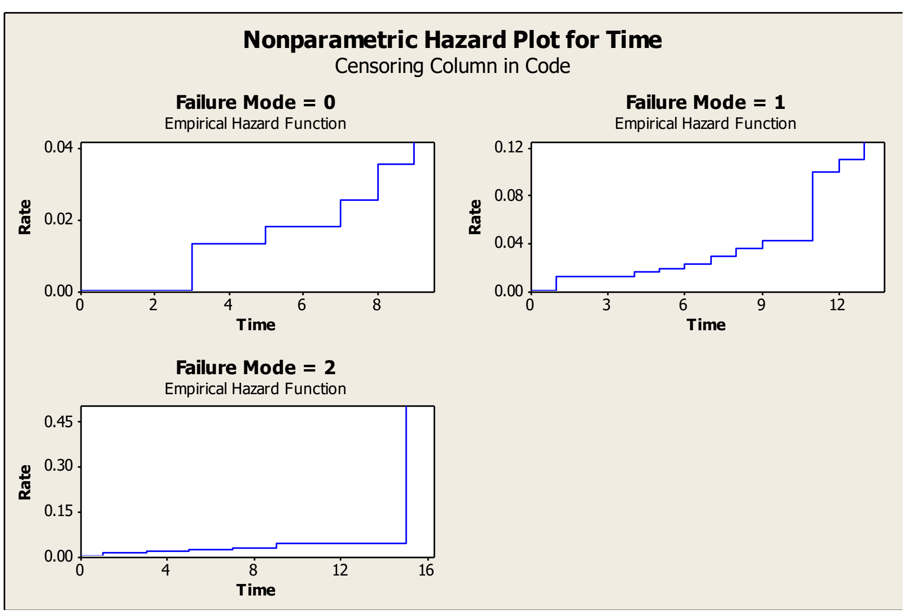
The above plots show the *survival probabilities for Patients having, respectively, none, one or two Comorbidities*. For example the probability of a patient with *No Comorbidities* of surviving six days in the Ventilator is 95%. If the patient has *one Comorbidity*, the probability is 80%, and if it has *two Comorbidities*, the probability is 75%. Chances of survival decrease with number.

<u>Nonparametric Hazard Plot for Time in Ventilator, considering Censoring</u>



The hazard plot provides the instantaneous failure probability (death) of a Ventilator patient. The use of this plot is more conceptual. Notice how the hazard rate grows slowly (to about 0.03) for the first seven days, then moves up faster until 0.07 in the eleventh day, and then abruptly to 0.1 and above, after that. No consideration for Comorbidities are taken here, but Censoring is taken.

<u>Nonparametric Hazard Plot for Time, considering Comorbidities (F.M. = 0,1,2)</u>



For No Comorbidities, hazard rate is practically zero up to the third day, then grows to 0.015, and slowly there after to about 0.025. *For one Comorbidity* it grows slowly to 0,04 for the first ten days, then increases to 0.08 in the 11th day. *For two Comorbidities* it is about 0.1in ten days.

Summarizing the comparison of survival graphs: they show the strong efffect of Comorbidities on both, survival probabilities and hazard rates, of the ICU Ventilator patients.

4.0 Non Parametric Survival Analysis: Second Case

We now *consider the Second Case*. Assume *we conduct a study in an ICU during twenty days, and compute Time in Ventilators for two different groups of its 90 patients*. One group consists of the *39 patients still alive* at the time the study is terminated. The second group consists of the *remaining* 51 patients that have *passed away*. Now, we *analyze both groups separately, without considering any Censoring*. The two groups represent two different populations.

We again use the *Kaplan-Meier* Distribution Free survival method, for the same reason given before: we *do not want to commit to any specific distribution*. Analysis Results are given below.

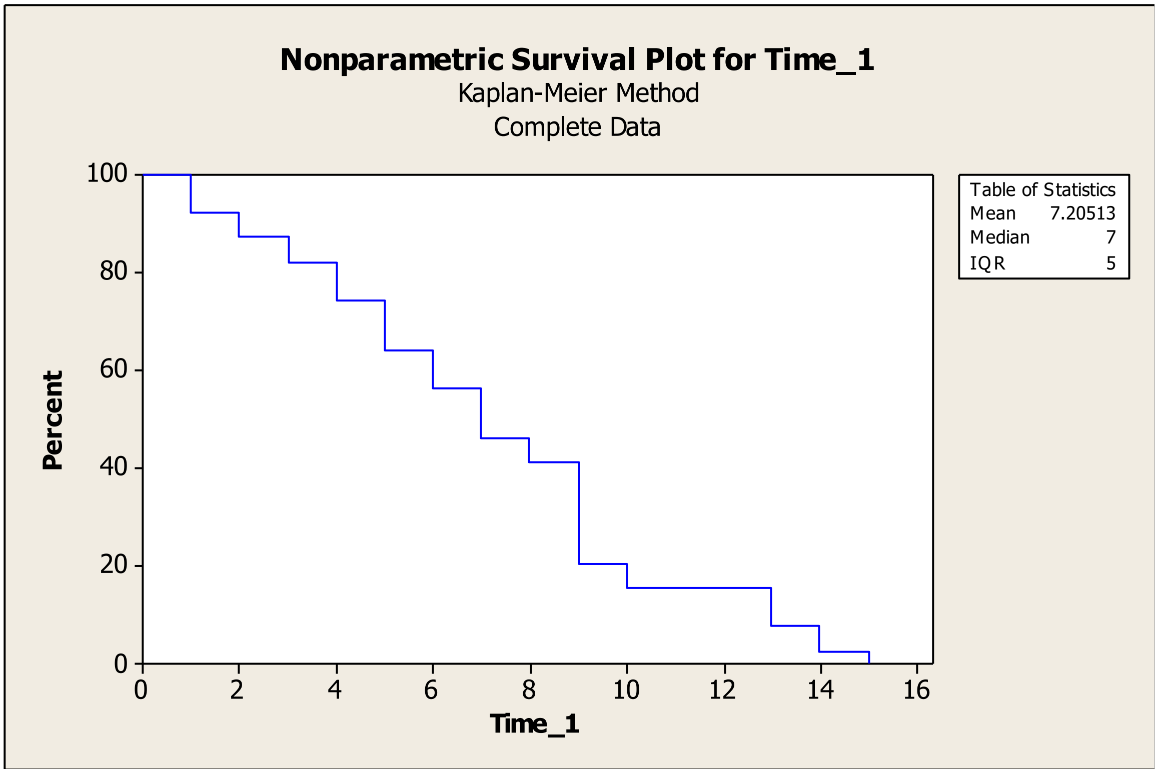
Group One (Ventilator patients that have improved/ recovered and have left the device):

```
Censoring Information  Count
Uncensored value          39

Nonparametric Estimates

Characteristics of Variable Time in Ventilator

            Standard   95.0% Normal CI
Mean(MTTF)     Error    Lower    Upper
   7.20513  0.612708  6.00424  8.40601 (Notice MTTF/leaving Ventilator, is smaller)

Median = 7; IQR = 5  Q1 = 4  Q3 = 9     (Non Parametric Descriptors)
```

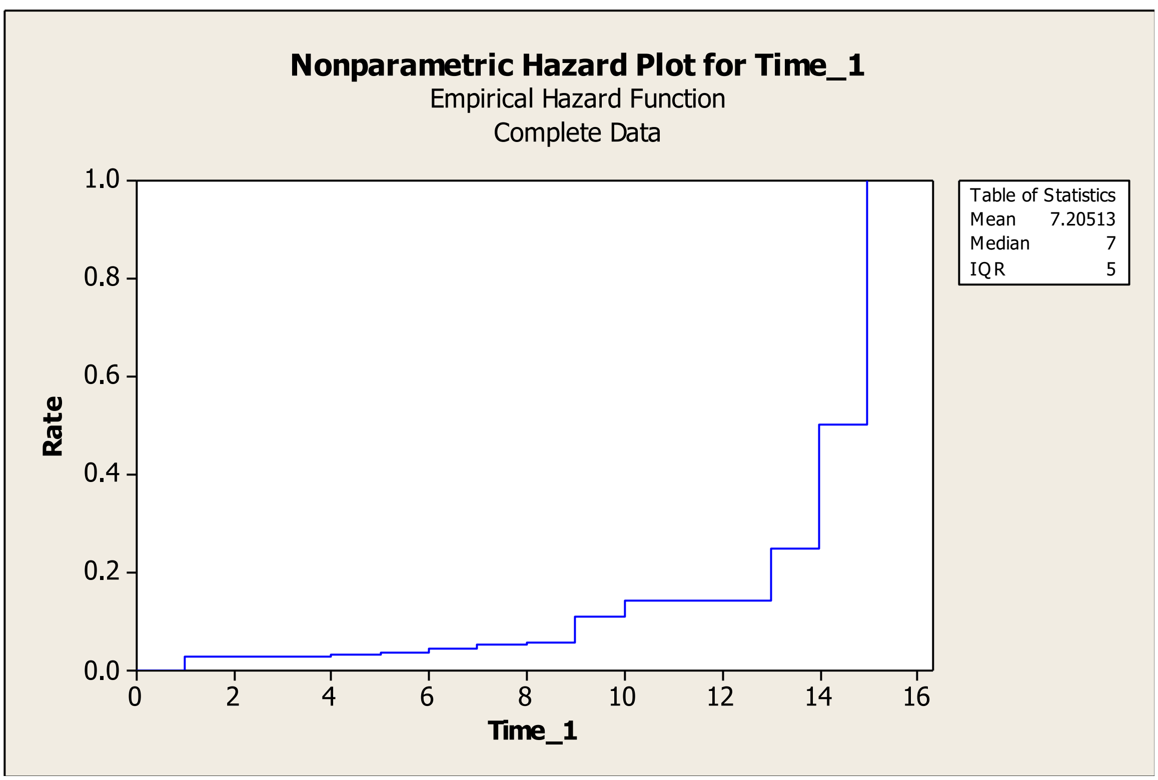
Table 3: Kaplan-Meier Estimates

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	95.0% Normal CI Upper
1	39	3	0.923077	0.0426692	0.839447	1.00000
2	36	2	0.871795	0.0535337	0.766871	0.97672
3	34	2	0.820513	0.0614507	0.700072	0.94095
4	32	3	0.743590	0.0699201	0.606549	0.88063
5	29	4	0.641026	0.0768134	0.490474	0.79158
6	25	3	0.564103	0.0794034	0.408475	0.71973
7	22	4	0.461538	0.0798268	0.305081	0.61800
8	18	2	0.410256	0.0787639	0.255882	0.56463
9	16	8	0.205128	0.0646590	0.078399	0.33186
10	8	2	0.153846	0.0577744	0.040610	0.26708
13	6	3	0.076923	0.0426692	0.000000	0.16055
14	3	2	0.025641	0.0253102	0.000000	0.07525
15	1	1	0.0000000	0.000000	0.00000	

These survival probabilities apply for Patients considered capable of recovery. No consideration is taken for Comorbidities. Further into this paper we develop Discrimination functions that help assess whether a Patient is expected or not to recover, given Age and Number of Comorbidities.



The same considerations as before are valid for these graphs; *interpretation differs*. For example, the probability that *a Patient will remain in a Ventilator* for six days or more, is about 60%



Group Two (Time in Ventilator for patients that have passed away):

Distribution Analysis: Time_1_1

```
Censoring Information  Count
Uncensored value          51
```

Nonparametric Estimates

Characteristics of Variable *Time in Ventilator*

Mean(MTTF)	Standard Error	95.0% Normal CI Lower	Upper	
5.37255	**0.454141**	**4.48245**	**6.26265**	**(Mean Sojourn Time prior to Death)**

Median = 5; IQR = 4 Q1 = 3 Q3 = 7   (Non Parametric Descriptors)

Table 4: Kaplan-Meier Estimates

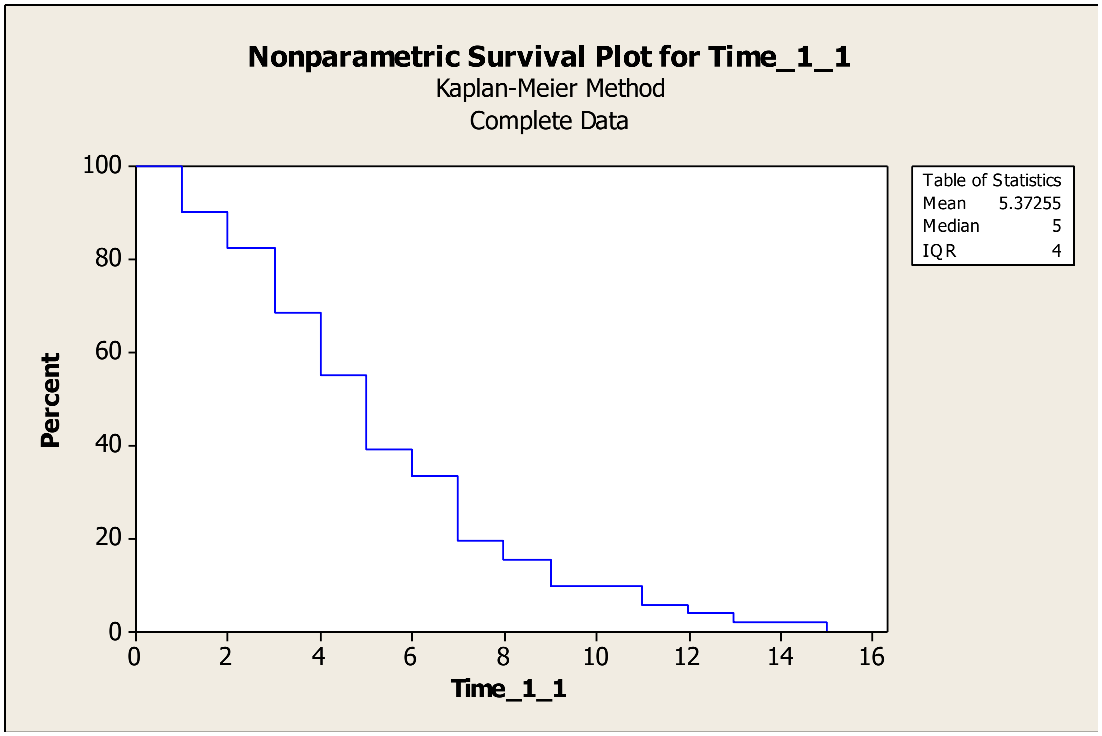
Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95.0% Normal CI Lower	Upper
1	51	5	0.901961	0.0416398	0.820348	0.983573
2	46	4	0.823529	0.0533815	0.718904	0.928155
3	42	7	0.686275	0.0649739	0.558928	0.813621
4	35	7	0.549020	0.0696767	0.412456	0.685583
5	28	8	0.392157	0.0683661	0.258162	0.526152
6	20	3	0.333333	0.0660098	0.203956	0.462710
7	17	7	0.196078	0.0555951	0.087114	0.305043
8	10	2	0.156863	0.0509242	0.057053	0.256672
9	8	3	0.098039	0.0416398	0.016427	0.179652
11	5	2	0.058824	0.0329478	0.000000	0.123400
12	3	1	0.039216	0.0271805	0.000000	0.092489
13	2	1	0.019608	0.0194147	0.000000	0.057660
15	1	1	0.000000	0.0000000	0.000000	0.000000

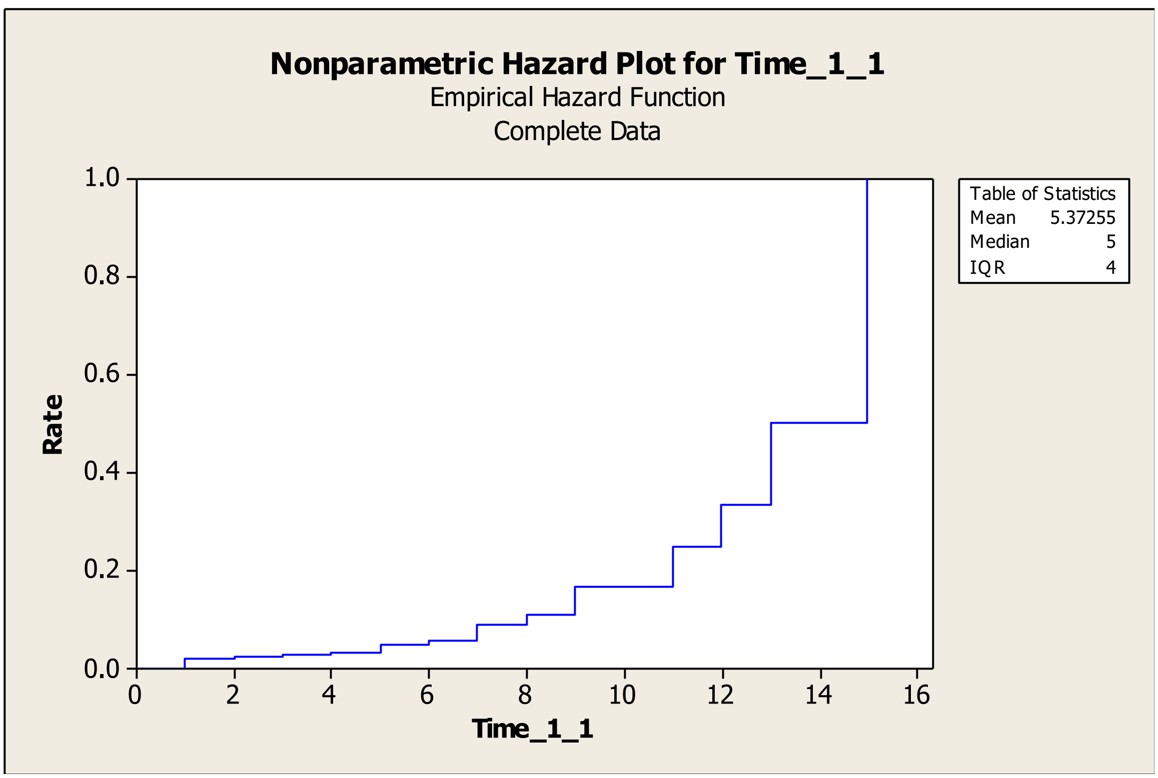
This survival table may apply to such Patients that have been assessed as too ill to survive in the Ventilator. Discriminant Functions to classify Patients, given Age and Number of Comorbidities into surviving or not in the ICU Ventilator, will be developed later in this paper.

An example of Table 4 interpretation is: the probability that a Patient will survive six days in the Ventilator is 0.333 (33.3%) and a 95% CI for such survival probability ranges from 20% to 46%.

In the unfortunate case of having to implementing Triage, where there are not enough Ventilator Units, they are all in use, and a young Patient arrives in need of one, such Tables may be used to decide to release currently occupying Patients with very low probabilities of survival.

The survival plots below, as before, are graphical interpretations of the above survival tables. Recall that these probabilities were obtained without consideration to Patient Comorbidities.





Interpretation of these survival and hazard rate plots are similar to the interpretations given in the previous section. Again, recall that these probabilities were obtained without any consideration to Patient Age or Number of Comorbidities.

We present below a summary table with three survival probabilities: (1) considering all 90 data points as either deaths, or *Censored* data (for Patients that have not died yet) assuming that they will perish at a later date than the termination of the study; (2) considering only *Censored* data, but now interpreted as Patient Sojourn Time to Recover in the Ventilator; and (3) considering only Time to Death in the Ventilator units.

Table 5: Comparison of the different Survival Probabilities

Fail. Time	Surv.Prob.ALL(*)	Surv.Prob.Surv.	Surv.Prob.Died
1	0.944444	0.923077	0.901961
2	0.898374	0.871795	0.823529
3	0.815629	0.820513	0.686275
4	0.730414	0.743590	0.549020
5	0.627900	0.641026	0.392157
6	0.586040	0.564103	0.333333
7	0.480853	0.461538	0.196078
8	0.446507	0.410256	0.156863
9	0.390693	0.205128	0.098039
11	0.319658	0.153846	0.058824
12	0.284141	0.076923	0.039216
13	0.248623	0.025641	0.019608
14	0.124311	0.000000	0.000000

(*) Considering those Ventilator patients not yet dying as "censored"

One usage of the above tables is to consider incoming Patients in all three possibilities: (1) not knowing what will actually happen; but considering that a Patient may die; (2) assuming that the Patient will surely recover; and (3) assuming that the Patient will surely die.

For example, the probability that a Patient will stay at least six days in the ICU Ventilator will be: for (1), 0.58 (or 58%); for (2), 0.56 (or 56%); for (3), 0.33 (or 33%).

In the next section we develop Discriminant Functions that can be used in classifying a Patient into likely to survive or to die, based upon Age and Number of Comorbidities.

5.0 Discriminant Analysis

In this section we test if *the two factors* considered (Patient Age *and* Number of *Comorbidities*) are significant (*have an impact in Patient Survival*). Then we use these two factors to *develop* a *Discriminant Function* whereby an incoming Patient's survivability (or lack thereof) could be assessed. We *first* develop the Discriminant Function using the *Minitab procedure*, with our data. We *then* develop a Discriminant Function using *Fisher* approach (as done in a paper on Principal Component Analysis mentioned in the introduction). We are considering *two groups: recovering* (previously called Censored) *and dying Patients*, to derive said Discriminant functions.

Minitab Discriminant Procedure (ALL Data): Recovered versus Age, No, Comorbidities

Predictors: Age, Comorbidities

Group	0	1
Count	39	51

Summary of classification

	True Group	
Put into Group	0	1
0	38	3
1	1	48
Total N	39	51
N correct	38	48
Proportion	0.974	0.941

N = 90;   N Correct = 86;

Proportion Correct = 0.956   (Compare with Fisher's Miss-Classification Probability)

Squared Distance Between Groups

	0	1
0	0.0000	**12.1580**
1	**12.1580**	0.0000

(Compare with Fisher's Mahalanobis Distance)

Linear Discriminant Function for Groups

	0	1
Constant	-12.068	-34.647
Age	0.565	0.936
Comorb	1.128	3.433

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
26**	0	1	0	3.198	0.474
			1	2.989	0.526
43**	1	0	0	1.936	0.965
			1	8.560	0.035
64**	1	0	0	6.812	0.785
			1	9.403	0.215
67**	1	0	0	1.995	0.799
			1	4.750	0.201

The above are the only four (out of 90) data points erroneously miss-classified.

Fisher Regression Discrimination Analysis (All Data):

Divided into Two Groups (Deaths, Survival/Recovery) using Age, Comorbidities

The regression equation is: DscGrps = - 2.77 + 0.0459 Age + 0.286 Comorb

Predictor	Coef	SE Coef	T	P
Constant	-2.7651	0.1984	-13.94	0.000
Age	**0.045914**	**0.003785**	**12.13**	**0.000**
Comorb	**0.28564**	**0.08278**	**3.45**	**0.001**

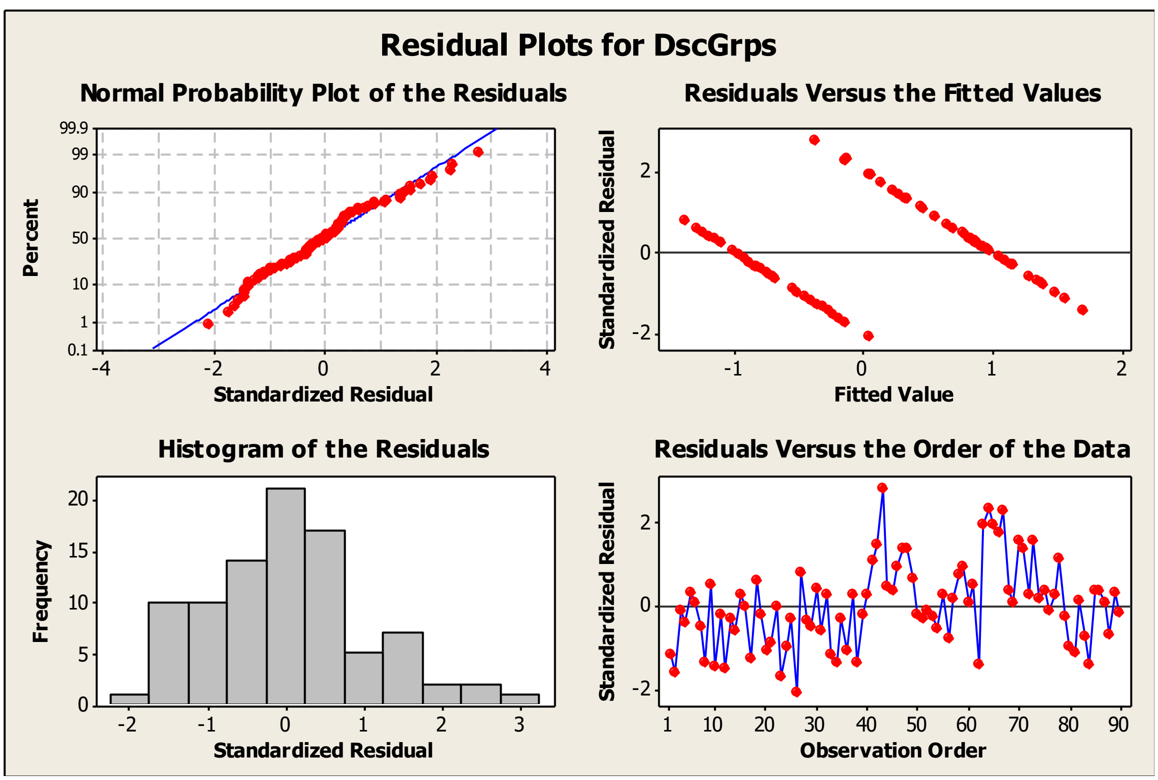
S = 0.500681 R-Sq = 75.3% R-Sq(adj) = 74.8% (Explains 75% of group differences)

Unusual Observations

Obs	Age	DscGrps	Fit	SE Fit	Residual	St Resid
26	55.0	-1.0000	0.0459	0.0545	-1.0459	-2.10R
43	52.0	1.0000	-0.3775	0.0865	1.3775	2.79R
64	45.0	1.0000	-0.1276	0.1307	1.1276	2.33R
67	51.0	1.0000	-0.1378	0.0597	1.1378	2.29R
90	84.0	1.0000	1.0917	0.1624	-0.0917	-0.19 X

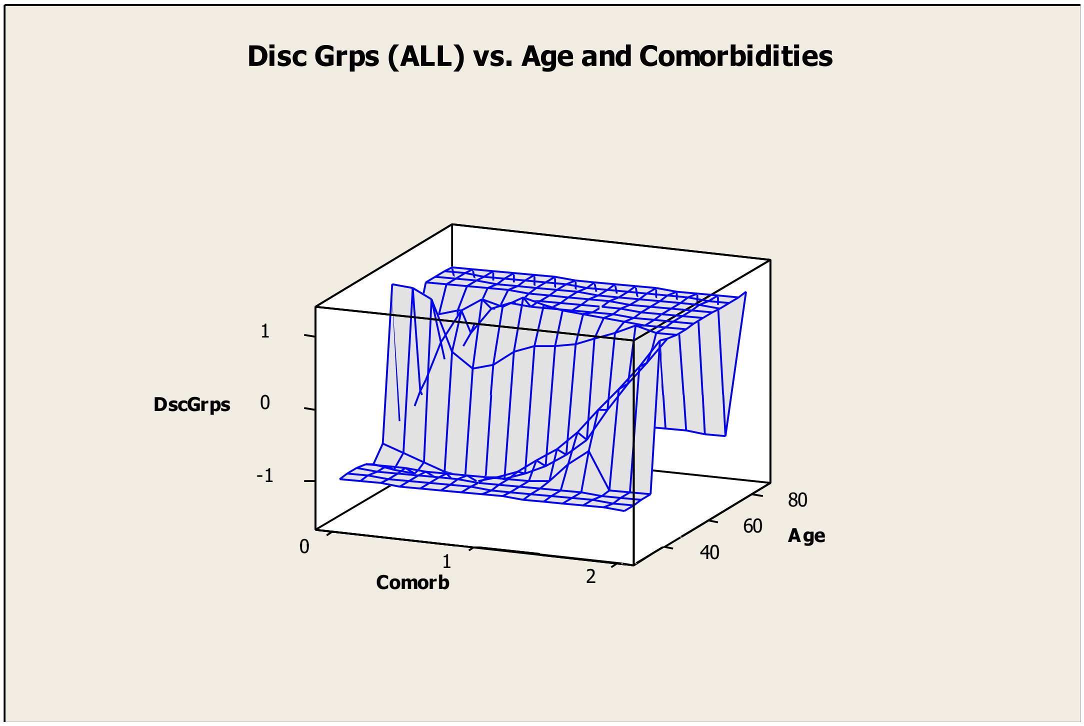
R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

Residual Plots for DscGrps (Deaths, Survivals)



Notice how the standardized residuals cluster about zero, as expected, and how this Discriminant Function explains 75% of the problem of classification of all observations into the two groups.

We present, below, a surface plot of the two groups (denoted -1, 1), separated by using said Discrimination Function. Notice how the two-group separation is well-delimited.



Mahalanobis Distance:

This Discrimination Function explains **75.3%** of the problem and is able to correctly classify most patients in their respective group. The **Mahalanobis Distance** that separates these two Extreme Groups (-1, 1), can be obtained in the following way:

For: n1= 51 (1); n2 = 39 (0);
Lambda ^{2} = n1*n2/(n1+n2) = 51*39/90 = 22.1
Dp^2 = [(n1+n2-2)/Lambda ^{2}]*[R^2 / (1- R^2)] =((90-2)/22.1)*(0.753/0.247)=12.14
Dp = Sqrt (Dp^2) = 3.48 => Prob (- ½ Dp) = Prob (-3.48 /2) = 0.0408

Comparison of results:

Disc. Function	Mahalanobis Distance	Prob. Misclassification	Factors
MINITAB	12.158	0.044	Age, Comorb
FISHER REGRESSION	12.14	0.041	Age, Comorb

Notice how the Mahalanobis Distance and Probability of Miss-Classification, between the two groups, established using both Discrimination Functions, is very similar for both Minitab and Fisher procedures. We can thus deduct two consequences from these results:

1. Factors age and No. Comorbidities are good metrics to differentiate the two groups, and
2. Both Discrimination Functions can be used to assess an incoming Patient's prognosis

6.0 Regression Analysis

We now perform failure analyses by implementing *regressions of Time to Exit Ventilator* units (by Death or Recovery). We use a regression approach[4] whereby the hazard function $\Lambda(*)$ can be written as a function of Time to Failure t, and a set of covariates Z, here represented by Patient Age and Number of Comorbidities, as well as a vector β of the coefficients of Z, to be estimated from the regression work. Said variables are defined:

$$Z = (z_1, z_2) = (\text{Age; No. Comorbidities}); \text{ and } \beta = (\beta_1, \beta_2) \text{ coefficients}$$

The Hazard Rate function is: $\Lambda(t; Z) = \mathrm{Exp}(\lambda_o + Z\beta)$, and the Likelihood function is:

$$\mathrm{Ln}\{\Lambda(t; Z)\} = \mathrm{Ln}\{\mathrm{Exp}(\lambda_o)\mathrm{Exp}(Z\beta)\} = \lambda_o + Z\beta = \lambda_o + z_1\beta_1 + z_2\beta_2$$

We use Ln(Time to Fail) as a Response in the regression v. Age and No. Comorbidities.

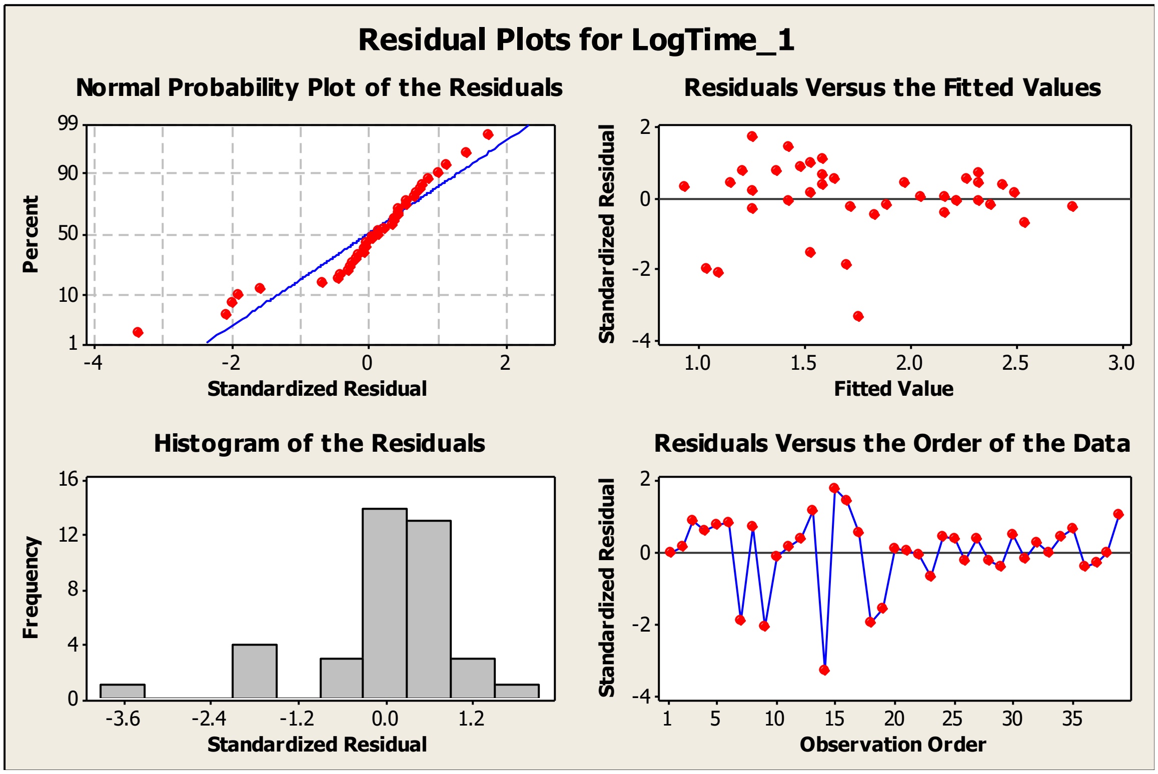
We implement the regression analysis using two Groups: Death & Recovery

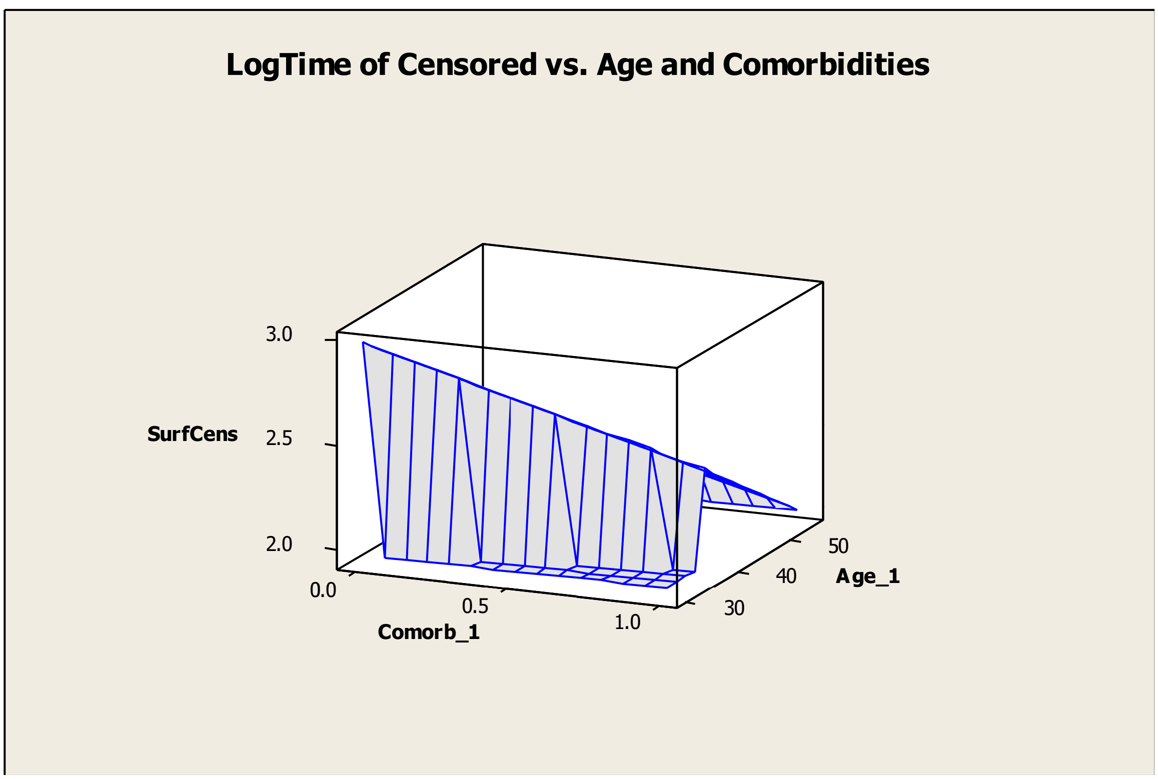
Regression Analysis: LogTime_1 v. Age_1, Comorb_1 for Survival/Recovery

```
The regression equation (for Survival; Code=0) is:

LogTime_1 = - 0.728 + 0.0551 Age_1 + 0.465 Comorb_1

Predictor      Coef   SE Coef      T      P
Constant    -0.7278    0.7539  -0.97  0.341
Age_1        0.0551   0.01910   2.88  0.007
Comorb_1     0.4649    0.2159   2.15  0.038

S = 0.548275 R-Sq = 45.3% R-Sq(adj)= 42.3%; (Regres.; Explains 42%)

Unusual Observations

Obs  Age_1  LogTime_1     Fit  SE Fit  Residual  St Resid
  9   33.0     0.0000  1.0900  0.1637   -1.0900    -2.08R
 14   45.0     0.0000  1.7510  0.1637   -1.7510    -3.35R

R denotes an observation with a large standardized residual.
```

For the *Group Survivals/Recovered*, factors *Age and Number of Comorbidities are statistically significant.* Regression explains 42% of the problem. *Factor contribution to LogTime to Recover (Ending their Ventilator stage) is positive* for both factors. Thence, as both Patient Age and Number of Comorbilities increase, the *Time to Recover* of said Patient also increases (which is consistent with data results and intuition).

[4] For more on the Proportional Hazards model see Kalbfleisch and Prentice in the Bibliography section.

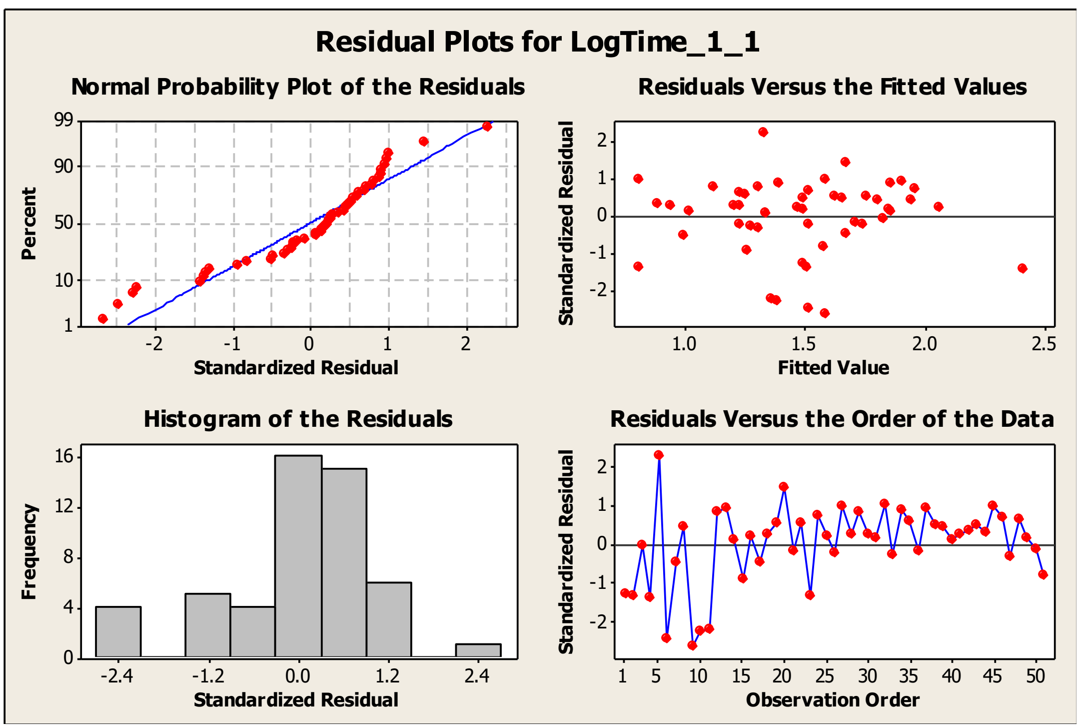




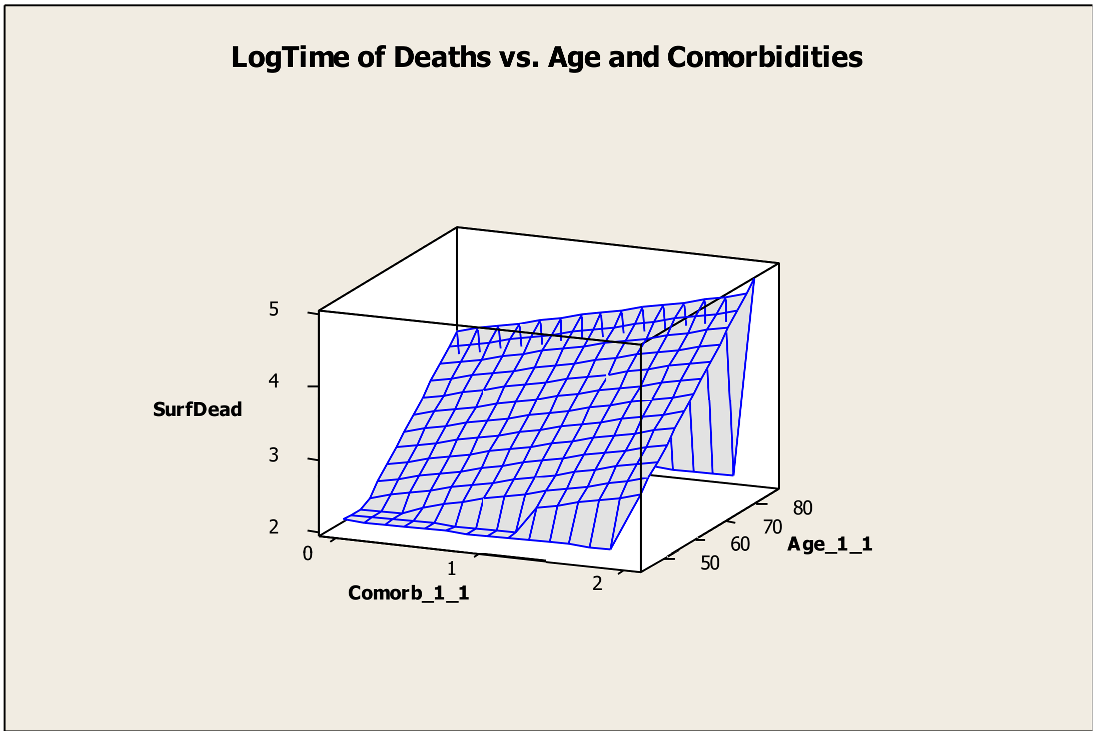
The above graph visualizes the Patient Age and Number of Comorbidities regions that make the function *Log Time of Censored/Survival* increase. Based on this Function we obtain (Table 6) estimates of the *Sojourn Time in a Ventilator* for Patients, given Ages and No. Comorbidities.

Regression Analysis: LogTime_1_1 versus Age_1_1, Comorb_1_1 (Death)

```
The regression equation (for Death; Code=1) is:

LogTime_1_1 = 3.76 - 0.0259 Age_1_1 - 0.371 Comorb_1_1

Predictor        Coef    SE Coef      T      P
Constant       3.7550     0.6590   5.70  0.000
Age_1_1     -0.025939   0.008685  -2.99  0.004
Comorb_1_1    -0.3712     0.1327  -2.80  0.007

S = 0.621788 R-Sq = 23.0% R-Sq(adj)=19.8% (Significant Regres.; Explains 20%)

Unusual Observations

Obs  Age_1_1  LogTime_1_1     Fit  SE Fit  Residual  St Resid
  4     52.0       1.6094  2.4062  0.2615   -0.7968     -1.41 X
  5     65.0       2.7081  1.3267  0.1287    1.3813      2.27R
  6     72.0       0.0000  1.5163  0.0967   -1.5163     -2.47R
  9     55.0       0.0000  1.5861  0.1668   -1.5861     -2.65R
 10     63.0       0.0000  1.3786  0.1327   -1.3786     -2.27R
 11     78.0       0.0000  1.3607  0.1176   -1.3607     -2.23R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```



Regression for the *Group Deaths* versus factors *Age and Number of Comorbidities is highly statistically significant* and explains 20% of the problem. Factors *Age and No. Comorbidities contribution to the LogTime to Death and End of Ventilator sojourn is negative* in both factors. *The higher the Age or Number of Comorbidities, the shorter the Time to Die in the Ventilator*.



The above graph visualizes the regions of Patient Age and Number of Comorbidities where the Log Time of Censored/Recovery function is greater. We will use this function as in the previous graph, to obtain expected values for Table 6.

We summarize the results of these regression analyses in Table 6, below. It shows the expected *Time to End* Patient *Ventilator sojourn*, *by Age and Number of Comorbidities*. Values under the Green headings correspond to estimations *calculated for group Recovery*. Those under Yellow headings correspond to estimations *calculated for group Deaths*.

Table 6: Comparison of the different Ventilator Sojourn Times

AgeEnd	T0ComR	T1ComR	T2Comr	T0ComD	T1ComD	T2ComD
35	3.3218	5.2883	8.419	17.3484	11.9712	8.26062
45	5.7632	9.1752	14.607	13.3899	9.2396	6.37574
55	9.9991	15.9188	25.343	10.3346	7.1313	4.92094
65	17.3484	27.6189	43.970	7.9765	5.5041	3.79810
75	30.0992	47.9184	76.287	6.1564	4.2482	2.93146
85	52.2218	83.1378	132.357	4.7517	3.2789	2.26257

For example, a 55 year old Patient, placed on Ventilator, will be expected to recover in 9.99 (10) days with Zero Comorbilities; in 15.9 days with One Comorbidity, and in 25.3 days, if with Two.

A Patient with the same 55 years age, will die after 10.3 days, with Zero Comorbidities; in 7.13 days, with One Comorbidity; and in 4.9 days, if with Two Comorbidities.

These are intuitive results. If this analysis is redone using actual hospital data, such results can be obtained and used in the fight against Covid-19

7.0 Discussion

In this paper we have *shown the use and power of* Survival Analysis methods for Covid-19,

A *data set has been created to illustrate* the implementation of *Survival Analysis* methods. These *results can be used (1) for research* (in the *identification and assessment of factors* that affect the hazard rate or probability of survival of a Ventilator Patient). They can also be used *(2) to help establish Patient hierarchy* for receiving medical treatment, if Covid-19 Pandemic becomes so critical that it is necessary *to implement Triage*.

Patients can be *evaluated using a Discriminant Function,* and their *probability of Recovery or Death in the Ventilator* can be *assessed*. In addition, their *Sojourn Time in the Ventilator*, based on their Age and Number of Comorbidities *can be estimated*. *Triage Decisions*, regarding Patient placement and treatment, may be taken *based upon such evaluations*.

8. Conclusions:

The main objective of the present analysis is *to provide* a detailed *example of the use of Survival Analysis, applied to* the study and assessment of *Covid-19 data*.

This framework can be used as a guide for the analysis of hospital data, or from any city, state or country. It can *include more factors* (metrics), as they become available to researchers. For example, instead of (or in addition to) Number of Comorbidities it can include each Comorbidity at its level, in the Patient (e.g. if considering Cancer, we give its Phase: I, II, III, IV)

Survival Analysis can also *help researchers generate new ideas for other types of analyses* that can be implemented *using Covid-19 data*.

Bibliography

Anderson, T.W. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons. New York. Second Edition. 1971.

Kalbfleisch, J. D. and R. L. Prentice. Statistical Analysis of Failure Time Data. John Wiley & Sons. New York. 1980.

Reliability Toolkit. Reliability Analysis Center. RAC.

Romeu, J. L. *A Comparative Study of Goodness-of-Fit Tests for Multivariate Normality*. Journal of Multivariate Analysis. V. 46, No. 2. August 1993. 309--334.

Romeu, J. L. 1978. *Analisis y Clasificacion de Fuentes de Informacion, mediante métodos estadisticos*. Revista de Ingenieria Civil. Vol. 4. Havana, Cuba. Available in the Internet: https://web.cortland.edu/romeu/DiscrAnalClasFuentesGvnDap.pdf

Romeu, J. L. RAC START. Vol. 9, No. 6: *Empirical Assessment of Normal and Lognormal Distribution Assumptions*. https://web.cortland.edu/matresearch/NormAssumSTART.pdf

Romeu, J. L. RAC START: Vol. 8, No. 2: *Statistical Assumptions of an Exponential Distribution.* https://web.cortland.edu/matresearch/ExpAssumSTART.pdf

Romeu, J. L. RAC START. Vol. 10, No. 3: *Empirical Assessment of the Weibull Distribution*. https://web.cortland.edu/matresearch/WeibAssumpSTART.pdf

Romeu, J. L. RAC START. Vol. 10, No. 7: *Reliability Estimations for Exponential Life*. https://web.cortland.edu/matresearch/RExpLifeSTART.pdf

Romeu, J. L. RAC START: Vol. 8, No. 2: *About Censored Data*. https://web.cortland.edu/matresearch/CensorDatSTART.pdf .

Romeu, J. L. QR&CII Tutorial Series, Vol. 1, No. 2*: Infant Mortality in Reliability Studies for the Weibull*. https://web.cortland.edu/matresearch/InfantMortStudyQR&CII.pdf

About the Author:

Jorge Luis Romeu retired Emeritus from the State University of New York. He was, for sixteen years, a Research Professor at Syracuse University and is currently an Adjunct Professor of Statistics. Romeu worked for many years as a Senior Research Engineer at the Reliability Analysis Center (RAC), an Air Force Information and Analysis Center operated by IIT Research Institute (IITRI). Romeu received seven Fulbright assignments: in Mexico (3), the Dominican Republic (2), Ecuador, and Colombia. He holds a doctorate in Statistics/O.R., is a C. Stat Fellow of the Royal Statistical Society (RSS), a Member of the American Statistical Society (ASA) and of the American Society for Quality (ASQ). He is a Past ASQ Regional Director and holds Reliability and Quality ASQ Certifications. Romeu created and directs the *Juarez Lincoln Marti Int'l. Ed. Project* (https://web.cortland.edu/matresearch/) dedicated to support higher education in Ibero-America.