**AA502 – Survival Analysis**

**Hurricane HW1**

Consider the following “hypothetical” situation. Several hurricanes struck the gulf area and resulted in severe casualty and property damage. One of the major defenses is to maintain and coordinate the pump operations during a critical 48 hour period or over 4 high tides. The Steering Committee of the Center for Risk Management is conducting a survival analysis for the pump stations in the gulf coast area. Data was collected from 771 pump stations. There are five potential failure conditions during a major hurricane hit recorded in the variable **reason**:

1. No failure
2. Flood – overflow or accumulation of an expanse of water that submerges the pump station.
3. Motor – mechanical failure.
4. Surge – onshore gush of water usually associated with levee or structural failure.
5. Jammed – accumulation of trash or landslide materials.

The factors that potentially influence the survivability of the pump stations are the following (not all pumps have each characteristic, but some characteristics are available through upgrade or maintenance):

* **Backup pump** – a redundant system used to protect the station from flooding when the main pump is not operating. (Upgrade available)
* **Bridge Crane** – allow vertical access to equipment and protecting materials. (Upgrade available)
* **Servo** – servomechanism is used to provide control of a desired operation through the Supervisory control and data acquisition (SCADA: Supervisory Control And Data Acquisition) systems. (Upgrade available)
* **Trash-rack cleaner** – protecting hydraulic structures against the inlet of debris, of vegetation, urban or industrial trash. (Upgrade available)
* **Elevation** – elevation of the pump station that can be altered by 1 foot by maintenance. (Maintenance available)
* **Slope** – surrounding ravine slope of the pump station.
* **Age** – difference between the installation and the current date.
* **H1 ~ H48** – pumping status during a 48 hour emergency reported by pump stations – accuracy of pump status not guaranteed to be error free.

As the Steering Committee of the Center for Risk Management, provide a report and set of recommendations summarizing the findings from the survival analysis performed on the data. The report should include the following information:

* Provide summary statistics for each of the types of pump station failure. What percentage of pumps survived the hurricane? What percentages of pumps are in each failure? What is their average survival time? Are these averages different?

| **reason** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **0** | 316 | 41.04 | 316 | 41.04 |
| **1** | 115 | 14.94 | 431 | 55.97 |
| **2** | 112 | 14.55 | 543 | 70.52 |
| **3** | 111 | 14.42 | 654 | 84.94 |
| **4** | 116 | 15.06 | 770 | 100.00 |

reason=0

| **Analysis Variable : hour** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 316 | 48.0000000 | 0 | 48.0000000 | 48.0000000 |

reason=1

| **Analysis Variable : hour** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 115 | 26.4434783 | 13.2633192 | 1.0000000 | 48.0000000 |

reason=2

| **Analysis Variable : hour** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 112 | 41.0446429 | 10.6003471 | 3.0000000 | 48.0000000 |

reason=3

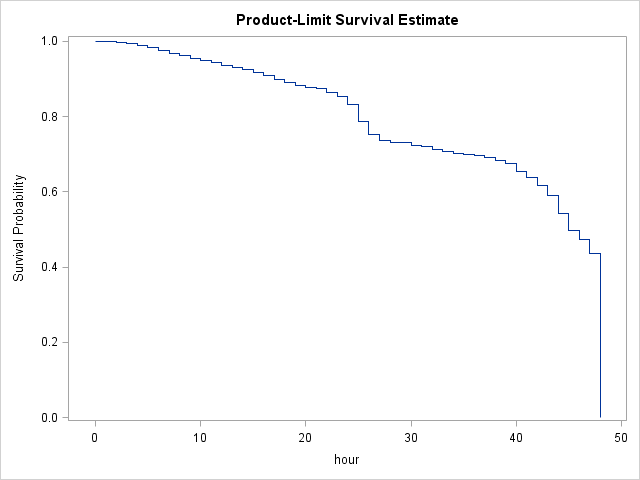
| **Analysis Variable : hour** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 111 | 38.8288288 | 10.8626749 | 2.0000000 | 48.0000000 |

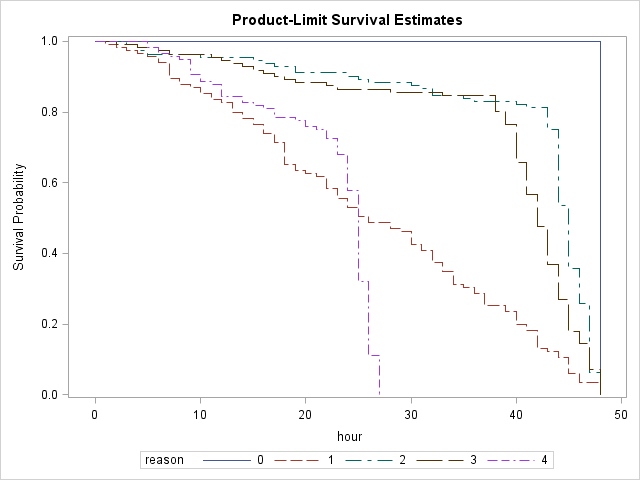
reason=4

| **Analysis Variable : hour** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 116 | 21.9396552 | 6.2572237 | 5.0000000 | 27.0000000 |

Reason 1 & 4 and Reason 2 & 3 have similar averages for when they failed.

* Plot two graphs (include these in the report):
  + The survival curve for all surviving pumps
  + The overlapping survival curves stratifying the data into reasons for failure.
  + Discuss any interesting things you find.



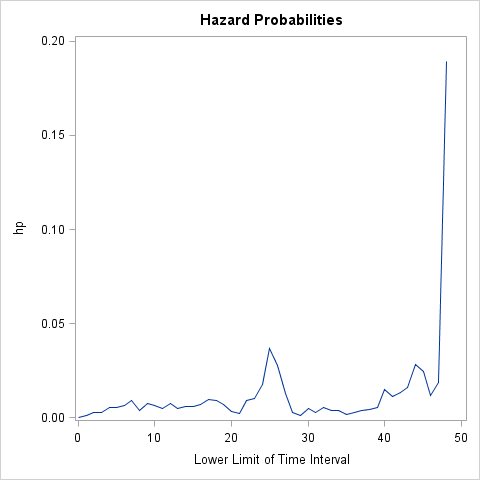


Pump Failure reason 2 and 3 have a similar curve. 1 and 4 share a similar curve until ~hour 20.

* Do the four major types of pump failures have similar survival curves? Use the DIFF=ALL option in the STRATA statement in SAS to get all pairwise tests.

All are different except 2-3 and 1-4.

| **Adjustment for Multiple Comparisons for the Wilcoxon Test** | | | | |
| --- | --- | --- | --- | --- |
| **Strata Comparison** | | **Chi-Square** | **p-Values** | |
| **reason** | **reason** | **Raw** | **Tukey-Kramer** |
| **0** | **1** | 606.3 | <.0001 | <.0001 |
| **0** | **2** | 290.9 | <.0001 | <.0001 |
| **0** | **3** | 355.1 | <.0001 | <.0001 |
| **0** | **4** | 741.7 | <.0001 | <.0001 |
| **1** | **2** | 42.7165 | <.0001 | <.0001 |
| **1** | **3** | 25.4795 | <.0001 | <.0001 |
| **1** | **4** | 5.6720 | 0.0172 | 0.1203 |
| **2** | **3** | 2.2305 | 0.1353 | 0.5667 |
| **2** | **4** | 78.5758 | <.0001 | <.0001 |
| **3** | **4** | 54.5754 | <.0001 | <.0001 |

* A coworker naively proposes to combine the failures into two groups, water-based (flood/surge) and mechanical (motor/trash). Do you agree with the proposal? Explain.
  + No. They have different survival curves. The groupings noted above should be used if you had to combine the failure types.
* Hazard probabilities are sometimes more intuitive than hazard rates. Unfortunately, SAS does not provide a graph of these by default. Take the following steps to create your own:
  + Use the METHOD=LIFE option in the PROC LIFETEST statement with a WIDTH=1.
  + In the PROC LIFETEST statement, use the OUTSURV=… option where the “…” is the name of a data set that will contain all of the needed information.
  + In this data set is a list of variables including **PDF** and **SURVIVAL**.
  + The hazard probability is the product of these two variables.
  + Plot the hazard probabilities for each failure type on one graph.
  + Discuss any interesting things you find.
    - There is a big spike around Hour 25.