Survival Analysis for the Army Corps of Engineers - Homework 3

Blue team 12

Sophee Li, Chris Nobblitt, John Pamplin, Jerry Liao

Executive Summary

The purpose of this report is to analyze the factors that influence the risk of pump failure due to motor and surge issues for the Army Corps of Engineers. We found that the Cox regression model adequately describes the pump failure time and that running a pump for 12 hours consecutively substantially increases failure rate. The concordance rate of our final Cox regression model is 0.77, meaning our model assigns the higher predicted hazard value to the subject that had the event earlier 77% of the time.

Results

During Hurricane Katrina, 41% of 770 pump stations survived during the 48-hour critical period. Being jammed by trash or landslide material was the most likely cause of pump stations failure (15.1%), followed by being flooded (14.9%), mechanical failure (14.5%), and structural damage (14.4%).

In order to identify the factors that influence mechanical failure and structural damage, we built an AFT model (Figure A1 to A4) and a Cox regression model with seven variables (backup pump, bridge crane, servomechanism, trash rack, elevation, slope, and age). Compared with the AFT model, whose underlying assumption of following certain distributions was violated, the Cox regression model describes our data more accurately. This is due to the fact that the Cox regression model does not require any distributional assumptions, unlike the AFT model.

The result of the Cox regression model is summarized in Table 1 below. The variable *Trashrack* and *Slope* appear to be statistically significant to a pump's probability of failure. The risk of a pump's failure is 0.34 times for a pump with an upgraded trash rack cleaner than the one without an upgrade. For each additional increase in age, the probability of failure is 1.13 times higher.

Table 1. Summary of the Cox Regression Model

	coef	**exp(coef)**	**se(coef)**	**z**	**Pr(>|z|)**
backup	-0.0934	0.9109	0.1368	-0.6820	0.4950
bridgecrane	-0.0330	0.9676	0.2037	-0.1620	0.8714
servo	0.2789	1.3217	0.1629	1.7120	0.0868
trashrack	-1.0669	0.3441	0.1479	-7.2150	0.0000
elevation	-0.0720	0.9305	0.0769	-0.9370	0.3488
slope	-0.1003	0.9046	0.0349	-2.8750	0.0040
age	0.1197	1.1271	0.0683	1.7520	0.0797

We tested if the seven variables in the Cox regression model are constant over time. After plotting each variable via ZPH plot, *Backup* (Figure 1), *Servo* (Figure 2), *Age* (Figure 3) and *Trashrack* (Figure 4) appear to be not constant over time, since the full line in each plot is not straight.

Figure 1. ZPH Plot for Backup	Figure 2. ZPH Plot for Servo

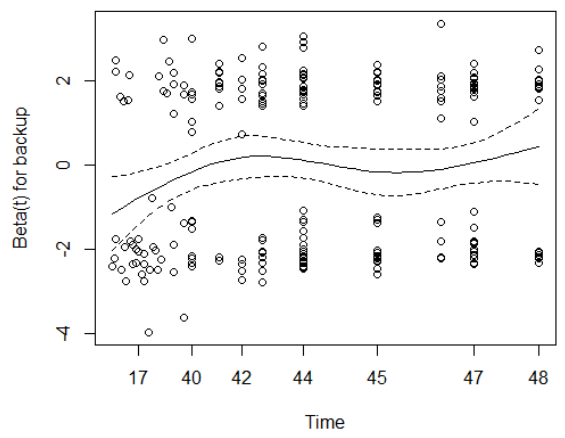
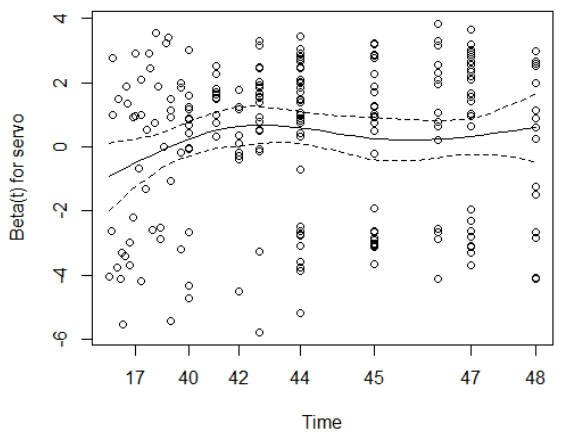
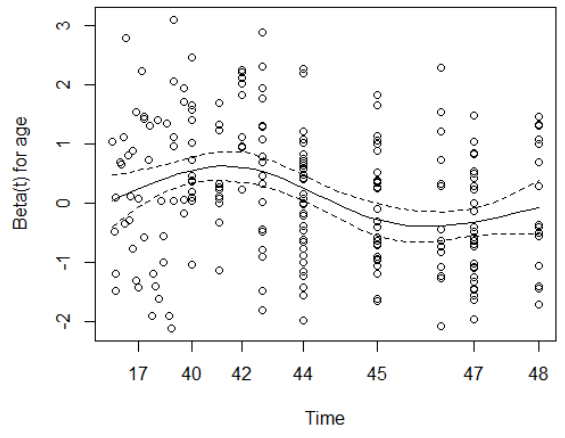
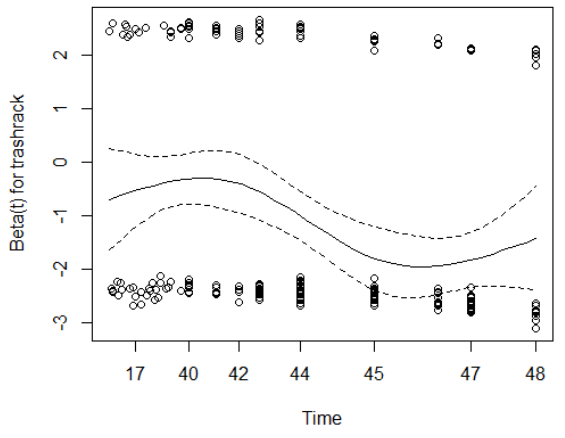

	

Figure 3. ZPH Plot for Age	**Figure 4. ZPH Plot for Trashrack**
	

The result of the Cox regression model with the new variable *Consecutive* is summarized in Table 2 below. The variable *Consecutive* describes whether the pump ever ran for 12 hours consecutively. We found out that the pumps running for consecutive 12 hours were 1.14 times more likely to fail than the ones did not (Table 2).

Table 2. Summary of the Cox Regression Model II

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
backup	0.0169	1.0171	0.0196	0.8630	0.3884
bridgecrane	-0.0592	0.9426	0.0293	-2.0190	0.0435
servo	0.2929	1.3403	0.0238	12.2900	0.0000
trashrack	-1.1637	0.3123	0.0213	-54.5080	0.0000
elevation	0.0314	1.0319	0.0109	2.8860	0.0039
slope	-0.1820	0.8336	0.0052	-34.9470	0.0000
age	0.1554	1.1682	0.0099	15.6380	0.0000
consecutive	0.1276	1.1362	0.0244	5.2380	0.0000

Conclusion

Our analysis helps the Army Corps of Engineers understand what factors affect the risk of pump failure. We found that bridge crane, servo, trashrack, elevation, slope, and age significantly influence the risk of pump failure. In addition, we confirmed that running pumps for 12 consecutive hours increases the risk of pump failure. We suggest occasionally shutting down pumps to prevent them from overloading. We would recommend an analysis of the length of time needed for a pump to “recover” after being used for 12 consecutive hours. This analysis would help to prevent pump failures by creating guidelines for the next flooding event.

Appendix

Figure A1. Exponential Distribution	Figure A2. Log-Logistic Distribution

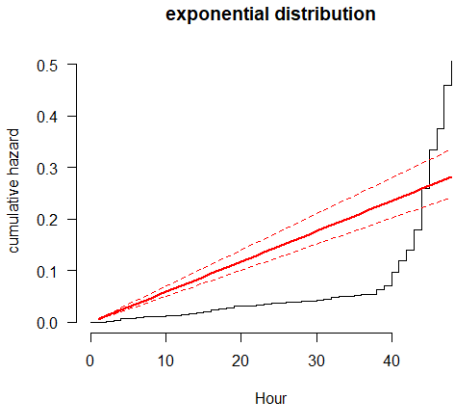
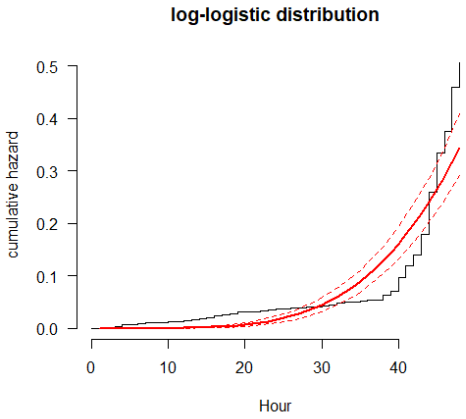

	

Figure A3. Lognormal Distribution	**Figure A4. Weibull Distribution**