MATH-449 - Biostatistics
EPFL, Spring 2025
Problem Set 6

1. A statistics student did an internship with a power company, where she was hired to analyze the time it took before cracks developed in the company's new turbine prototype. 41 turbines were observed in a testing facility for two months, where engineers had carefully recorded the time it took before noticeable cracks were discovered. The power company had installed machines that could make turbines rotate to simulate the rotation they could experience in a real-world environment.

   The following outcomes were recorded:

   - Some of the turbines developed cracks before the two months were over.
   - Some turbines were observed for the whole two months without any cracks.
   - For a significant number of turbines, the machines that enforced the rotation stopped working during the study. The power company didn't have the resources to repair these machines, making the turbines they rotated unobserved.

   The power company was interested in the time it took before cracks developed if the rotation enforcing machines did not stop. The turbines whose machines stopped were considered censored. The time it took for the machines to fail was thought to be unrelated to the time it took before cracks developed so that the observed (non-censored) turbines were representative of all turbines.

   a) Classify the above outcomes using the variables $\tilde{T}_i$ and D_i from the lectures.

   Being familiar with survival analysis, the student calculated the Kaplan-Meier curve, which estimates the survival probability as a function of t. The estimate along with approximate 95% confidence intervals (for each fixed t), $\hat{S}(t) \pm 1.96\hat{\sigma}(t)$, is plotted in Figure 1.

   b) Based on the plot, find the probability of a turbine being crack-free after 30 days, with a 95% confidence interval. Use the plot to estimate the 10th percentile of the survival times, along with a 95% confidence interval.

   The Kaplan-Meier estimator estimates the true survival probability $P(T > t)$ as long as the censoring is independent.

   c) Given the information provided thus far in this example, argue that the censoring is independent.

   Challenge: After talking with some of the engineers, the student learned that the machines provided different rotational speeds to the turbines. She reasoned that the machines that provided higher rotational speed: 1) could make cracks appear faster due to increased stress on the turbines, and 2) were more likely to stop working during the study (thus leading to censoring) due to increased stress on the machines. She learned that the machines could broadly be categorised into two groups; those that provided fast rotational speed and those that provided slow rotational speed.

   After thinking about the problem for a bit, she realised that the result from b) could provide a misleading picture of the survival probability. However, she also found that she could use the extra information about the machines' rotation speeds to improve her statistical analysis.

   d) Can you guess what she did?[2]

[2] Hint: what do we know about the censoring?

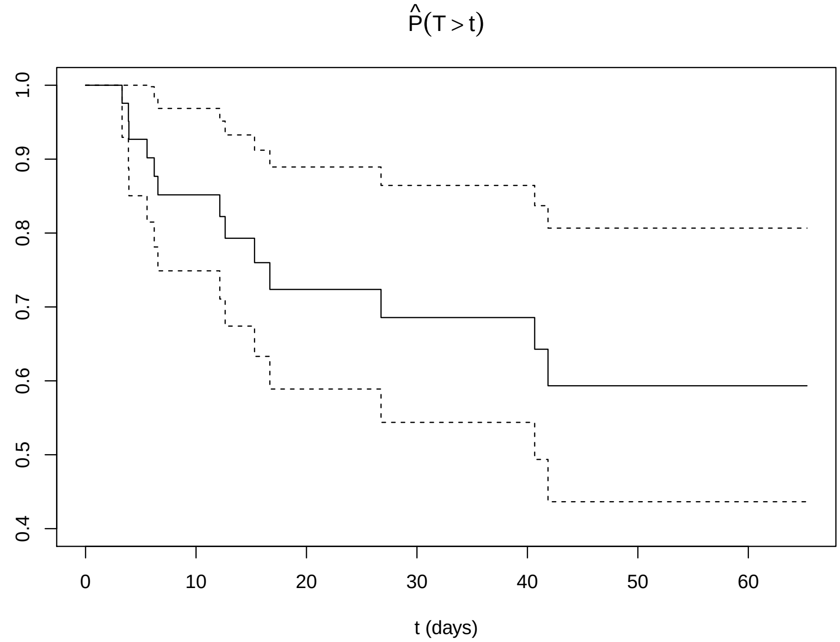

Figure 1:

2. Let T^1 and T^2 be two independent survival times with respective hazards α^1 and α^2.
   a) Show that $S = \min(T^1, T^2)$ has hazard $\alpha^1 + \alpha^2$.
   b) **Challenge:** Show that $P(T^1 < T^2 | S = t) = \frac{\alpha^1(t)}{\alpha^1(t)+\alpha^2(t)}$. [1]

3. (Exercise 3.1 from ABG 2008) The data in the table are from Freireich et al. (1963) and show the result of a study where children with leukemia are treated with a drug (6-MP) to prevent relapse, and where this treatment is compared with placebo. The numbers in the table are remission lengths in weeks; a '*' indicates a censored observation.

Placebo	1	1	2	2	3	4	4	5	5	8	8
	8	8	11	11	12	12	15	17	22	23	
6-MP	6	6	6	6*	7	9*	10	10*	11*	13	16
	17*	19*	20*	22	23	25*	32*	32*	34*	35*	

   a) Compute the Nelson-Aalen estimates for the 6-MP group and for the placebo group. [2]
   b) Plot both Nelson-Aalen estimates in the same figure. What can you learn from the plots?
   c) Calculate the Kaplan-Meier estimate for both groups, and plot the results.
   d) Estimate the probability of not having a remission the first ten weeks in both groups.

[1] Hint: consider the limit $\lim_{h \to 0+} P(T^1 < T^2 | t \leq S < t + h)$ and use the result from a).

[2] So far we have assumed absolutely continuous survival times, which implies that event times will not be tied. Note that some survival times are tied in the data set below. If T_j is an event times with d_j ties you may use the estimator $\Delta \hat{H}(T_j) = \sum_{i=0}^{d_j-1} \frac{1}{Z(T_j)-i}$, and set $\hat{H}(t) = \sum_{T_j \leq t} \Delta \hat{H}(T_j)$.