

Thorotrast Cancer Data Analysis

Aaron Moose

1 Data Description

Thorotrast was an x-ray imaging agent that was used between the 1920's through the 1950's. While being used, many people were diagnosed with some form of cancer. This analysis will be to determine what effect, if any, does Thorotrast have on expected survival. In addition, it will try to determine if there is an association between cancer related cause of death and Thorotrast exposure.

First, there is some modifications that will be done to the data set. There are two types of censoring when an individual is alive: censored at the end of the study, and censored as alive at some point before the end of the study. I will combine these into a single factor that will just represent censoring of any sort. Then, I will change the dosage into three levels: a dosage of 0mg, dosage of 0mg - 20mg, and a dosage greater than 20mg. Next, I removed individuals that died a violent death or are still alive. Lastly, I removed individuals that were not at least 40 years old at time of death. One reason for this is because the survival curves were crossing frequently before that point, but the main reason is that one group (males with a dosage larger than 20mg) didn't have anyone that died/censored before then.

2 Model Description

For the i^{th} individual, let T_i denote the death/censored time. Let $T_i \sim \text{Weibull}(\gamma, \lambda_i)$, with

$$\ln(\lambda_i) = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_5 z_{i5} + \varepsilon_i.$$

We are assuming that the shape parameter γ is a known constant while the scale parameter λ_i changes depending on the group of interest. The covariates are just indicators for the gender, the dosage levels, and the interactions between these two factors. The β_0 represents males that have a dosage level of 0 mg of Thorotrast.

3 Results

As stated before, I am assuming that the survival times follow a Weibull distribution. Based off of this assumption, the following is a table for the mean, standard error, and 95% confidence interval of every group.

Table 1: Numerical Summary

	Mean Survival Time	Std Error	Lower 95% CI	Upper 95% CI
Male, Dose 0	66.45	0.01	65.43	67.48
Male, Dose 1	62.30	0.01	61.15	63.48
Male, Dose 2	59.11	0.02	57.15	61.13
Female, Dose 0	72.76	0.01	71.59	73.94
Female, Dose 1	64.40	0.01	66.84	69.62
Female, Dose 2	62.61	0.02	61.95	67.60

So between genders, the overall mean survival times for males are less than the females that were given the same dosage level. This indicates that women do in fact live longer than males. The confidence intervals for males do not overlap although the males with dosage levels of 1 or 2 are close to each other. The female confidence intervals for a dosage level of 0 do not overlap with the other two intervals, but levels 1 and 2 do. This means that there is not a significant difference between the mean survival times of these two groups.

For the second question of the association between cancer and Thorotrast, it will be answered through a proportion test. So looking at whether an individual was exposed to Thorotrast or not and the number of people that have died from cancer, we can compute a proportion from each group that have cancer.

Table 2: Contingency table

	Total	Cancer Proportion
Control	1298	0.200
Thorotrast	793	0.322

Above is a contingency table detailing the total number in each group as well as the proportion in the groups that have died from cancer. Using a 2-sample test, the p-value obtained is 5.948×10^{-10} indicating that there is a significant difference between the proportions. In fact, the 95% confidence interval produced was $(-0.1614, -0.0811)$ which shows that the no exposure group had a small proportion having cancer. So there is some credibility to the claim of association between cancer and Thorotrast.