DTU

Project 2: Survival data

Introduction

This project studies the effect of **azidothymidine (AZT)** on the development of **AIDS** in patients with HIV.

Analysis of Binary Dataset

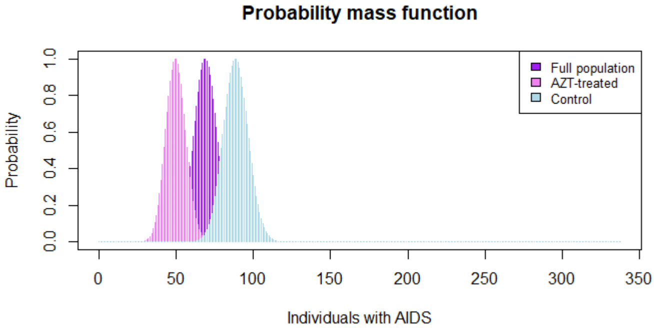
Binary dataset from a randomized study with two groups (AZT=yes vs. AZT=no) and two possible outcomes (AIDS=yes vs. AIDS=no)

AZT	AIDS	Total
Yes	25	170
No	44	168

AIDS	AZT	Control	Total
Yes	x=25	y=44	t=69
No	m-x=145	n-y=124	N-t=269
Total	m=170	n=168	N=338

Binomial Distribution

We fit binomial models on the full population and separately on the two groups and find that they are different.



Chi-squared test on proportions p-value: 0.012

Model of two binomial proportions

Using the log-odds ratio as difference parameter:

$$\beta_1 = log(\frac{\frac{prop_{AZT}}{1-prop_{AZT}}}{\frac{prop_{ct}}{1-prop_{ct}}}) \tag{1}$$

We built a model that accounts for the groups and optimized its likelihood function:

$$L(\beta_0, \beta_1) = e^{\beta_0 x} e^{\beta_1 (x+y)} (1 + e^{\beta_0 + \beta_1})^{-m} (1 + e^{\beta_1})^{-\beta_1} \tag{2}$$

Finding MLEs at:

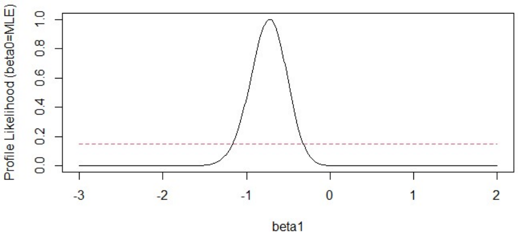
Parameter	MLE	SE
β_0	-1.0361	0.1754
β_1	-0.7218	0.2787

Model of two binomial proportions

We calculated the 95% Wald confidence intervals for this model:

Parameter	Lower CI	MLE	Upper CI
β_0	-1.38	-1.0361	-0.69
β_1	-1.27	-0.7218	-0.18

And also checked the interval in the profile likelihood plot:



The interval does not include 0, thus the groups are different.

Logistic Regression Model

We fit a General Linear Model with a logistic link function:

$$\log(L(\beta_0, \beta_1)) = \log \frac{y_i}{1 - y_i} = \beta_0 + x_i\beta_1 \tag{3}$$

And find parameter estimates and their CIs at:

Parameter	Estimate	SE	95% CI
β_0	-1.0361	0.1755	[-1.39; -0.70]
β_1	-0.7218	0.2787	[-1.27; -0.18]

Odds ratio

We calculate the odds ratio of the effect on AZT on AIDS:

$$OR = \frac{x/(m-x)}{y/(n-y)} = 0.4859 \tag{4}$$

And find its confidence intervals:

$$CI_{OR} = [0.2815, 0.8387] \tag{5}$$

The interval does not contain 1, thus the groups have different odds of developing AIDS. More specifically, the chances of AZT-treated HIV patients developing AIDS is 28-84% of those in the control.

Hypothesis Testing

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

- Likelihood ratio test

$$P\left(\chi_1^2 > -2 \log \frac{L(0)}{L(\beta_1)}\right) = 0.0085$$

- Wald test

$$P\left(z > \left|\frac{\beta_1}{SE_{\beta_1}}\right|\right) = 0.0096$$

- Score test

$$P\left(\chi_1^2 > S(\beta)^T I(\beta)^{-1} S(\beta)\right) = 0.0088$$

Analysis of survival time dataset

Survival time dataset from a double-blind, placebo-controlled trial with two groups (old treatment vs. new treatment)

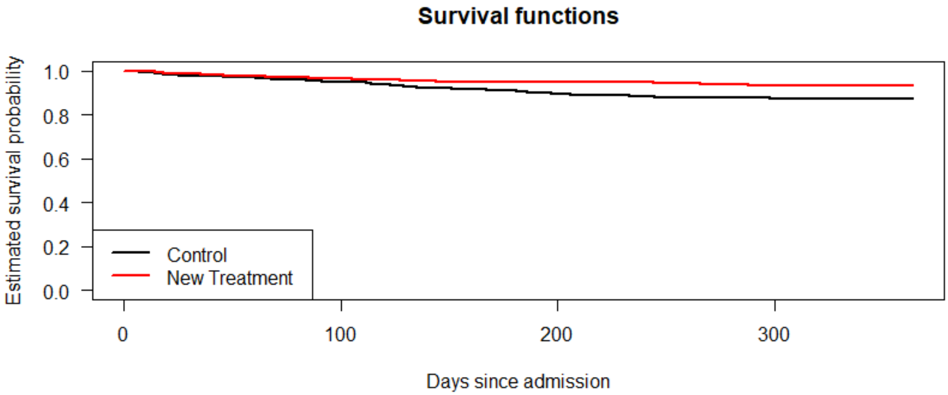
Variable	Meaning
`id`	Identification Code
`time`	Time to AIDS diagnosis or Death
`event`	Indicator for AIDS or Death
`tx`	Treatment indicator
`sex`	Sex
`raceth`	Race/Ethnicity
`karnof`	Karnofsky Performance Scale
`cd4`	Baseline CD4 count
`age`	Age at Enrollment

Descriptive Statistics

- 1151 Participants (577 in control and 574 in treatment group)
- 96 Patients with 0.083 proportion died or got AIDS in both groups
- 33 Patients with 0.058 proportion died or got AIDS in treatment group
- 63 Patients with 0.109 proportion died or got AIDS in control group
- Median follow-up time in control group was 251 days
- Median follow-up time in treatment group was 263 days

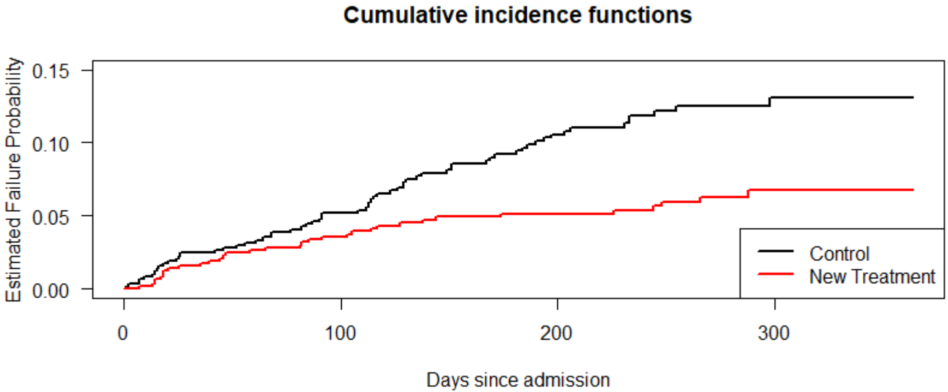
An important feature of this dataset is that some data is censored: we consider the probability density function for non-censored data and the cumulative probability function for the censored data.

Kaplan-Meier Survival Functions



Log-rank test shows curves are different with p-value 0.001

Cumulative Incidence Functions



Exponential Models 1 + 2 - Groups together and separated

We fitted exponential models to the full population and accounting for groups and optimized their likelihood functions:

$$L_1(\beta) = \frac{\beta e^{-\beta x_1}}{1 - e^{\beta x_0}} \tag{6}$$

$$L(\beta_0, \beta_1) = L_1(\beta_0) L_1(\beta_1) \tag{7}$$

To find the following MLEs and AICs:

Model	Parameters (1/rate)	AIC
Groups together	2759.8	1715
Groups separated	2052.1, 4017.2	1706

Likelihood ratio test proves a significant difference between the two, with p-value = 0.0008

Exponential Model 3 - Single parameter for treatment effect

We also build a model where a single parameter represents the difference between the groups (the treatment effect):

$$L(\beta_0, \beta_1) = L_1(\beta_0)L_1(\beta_0 + \beta_1) \tag{8}$$

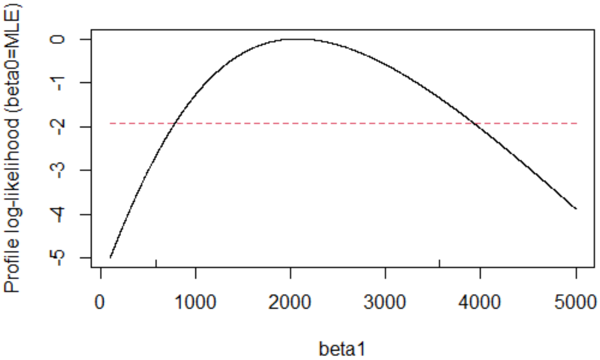
And find MLEs and Wald CIs:

Parameter	Estimate	Sd. error	95% CI
β_0	2047.5	257.96	...
β_1	2072.2	762.13	[578.5; 3566.0]

With an AIC of 1706, the likelihood ratio test proves there is no difference between this model and Model 2.

Exponential Model 3 - Single parameter for treatment effect

Looking at the CIs and the profile log-likelihood, we see that β_1 is different than 0 and conclude that there is a difference between the groups:



Parametric Survival Models - tx + cd4

Finally, we fitted a series of parametric survival models - containing treatment (tx) and CD4 count (cd4) as explanatory variables:

Model	Intercept - β_0	tx - β_1	cd4 - β_2	log(scale)	AIC
Exponential	6.71	0.66	0.01		1645.8
Weibull	7.05	0.84	0.02	0.24	1640.7
Log-Logistic	6.83	0.84	0.02	0.20	1639.7

And found the log-logistic was the best fit:

$$\log(T) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \sigma\epsilon \quad \epsilon \sim Logistic(0, 1) \tag{9}$$

The 95% Wald CIs of the estimated parameters were: $\beta_1 = [0.27 - 1.41]$ and $\beta_2 = [0.01 - 0.03]$

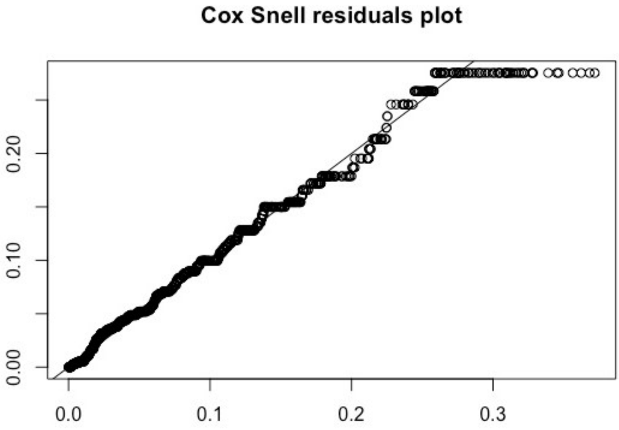
Parametric Survival Models - tx + cd4

We estimated the time ratio of the treatment effect and of an increased cd4 count effect:

	Time Ratio	2.5 CI	97.5 CI
tx	2.32	1.32	4.10
cd4 + 50	2.83	1.96	4.09

Both recieving treatment and increasing the cd4 count by 50 generates a 2.5x fold in the survival time.

Parametric Survival Models - Goodness of fit



Parametric Survival Models - Graphical Representation