# Survival Data Analysis

Nina Shenker-Tauris

02418 Statistical Modelling: Theory and practice
Jan Kloppenborg Møller

Technical University of Denmark
Department of Applied Mathematics and Computer Science

Winter Exam 2020

# Overview

## Analysis of Binary Data

The first part of the assignment deals with the study of the effect of AZT on AIDS.
Methods of analysis:

- Binomial distribution of all data
- Binomial distribution of treatment groups
- Parameter estimation of $p_0$ and $p_1$
- Profile likelihood of AZT-yes coefficient
- Logistic regression model

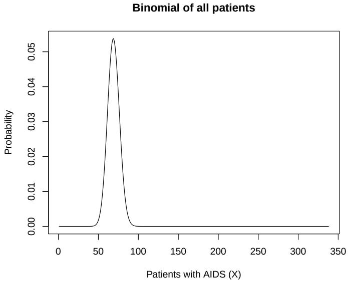
## Analysis of the Survival Time Data

The second part of the assignment deals survival time, where we want to see whether there is a difference for the two treatment groups
Methods of analysis:

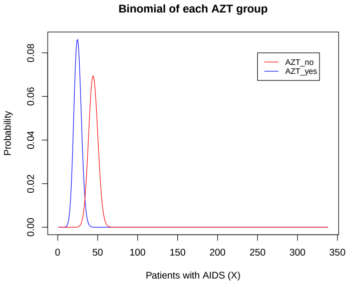
- AIDS/healthy statistics
- Survival distribution by treatment
- Cumulative incidence function
- Log-rank test
- Parametric survival model

# Binomial distribution of data

All data (AZT-yes and AZT-no)




Divided into AZT-yes and AZT-no




| 2-sample test for equality of proportions | | | | | |
|---|---|---|---|---|---|
| X-squared | df | p-value | 95% CI | prop 1 | prop 2 |
| 6.171 | 1 | 0.01299 | -0.21,-0.024 | 0.15 | 0.26 |

## Conclusion

P-value (0.01299) $<$ 0.05, therefore we reject the null hypothesis, i.e. there is a significant difference between the groups

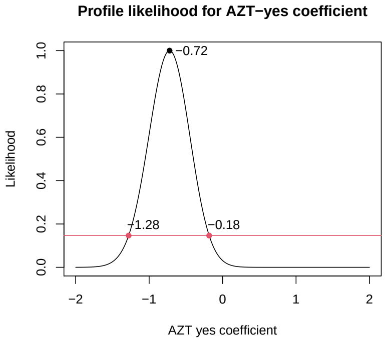
# Parameter Estimation

Estimate parameters in the model ($p_0$ probability of AIDS in control group, $p_1$ probability of AIDS in treatment group)

$$p_0 = \frac{e^{\beta_0}}{1 + e^{\beta_0}} \tag{1}$$

$$p_1 = \frac{e^{\beta_0+\beta_1}}{1 + e^{\beta_0+\beta 1}} \tag{2}$$

- $\beta_1$ = -0.7217; Therefore there is a negative correlation when given the treatment i.e. when given AZT your log odds of having AIDS are -0.721.
- Projecting back into probability scale, $p_0$ = 0.261, $p_1$ = 0.147, therefore, when given AZT you have a lower probability of having AIDS.




Confidence intervals using profiled log-likelihood: 95% CI [-1.279, -0.183]

# Logistics Regression Model

Logistic regression model for the binary outcome AIDS="yes" versus AIDS="no" with the explanatory variable treatment with AZT (yes, no)

$$logit < -glm(cbind(data\$AIDS\_yes, data\$n - data\$AIDS\_yes) \sim data\$AZT, data = data, family = "binomial") \quad (3)$$

Odds ratio for the effect of AZT on AIDS:

- GLM outputs log odds; if coefficient is positive, the effect of this AIDS is positive and vice versa.
- The odds of having AIDS with AZT: 0.147 / (1 - 0.147) = 0.172.
- The odd of having AIDS without AZT 0.261/(1-.261) 0.353.
- Computing a odd ratio we find that the odds of having AIDS is 2.05 times as high without immediate AZT treatment.

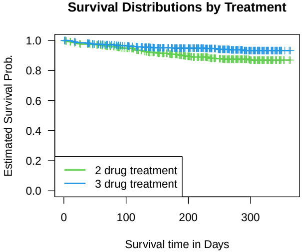
| Odds ratio 95 % CI | | |
|---|---|---|
| | 2.5 % | 97.5 % |
| (Intercept) | 0.2515723 | 0.5004942 |
| AZT Yes | 0.2813743 | 0.8390689 |

## Conclusion

GLM function supports earlier parameter estimation, in concluding that AZT has a significant effect on getting AIDS

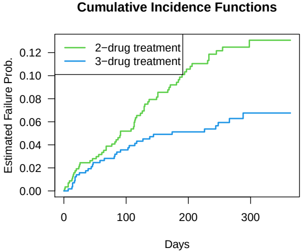
# Analysis of Survival Time Data

| Healthy vs AIDS/death | | | |
|---|---|---|---|
| | Healthy | AIDS | Probability of AIDS |
| Two-drug treatment | 514 | 63 | 0.11 |
| Three-drug treatment | 541 | 33 | 0.06 |
| Total | 1055 | 99 | 0.08 |




## Conclusion

Log-rank Test: This tests if the survival curves of the two treatment groups are the same. Result was p= 0.001, therefore it is significant ($p < 0.5$). The curves are indeed NOT the same meaning that the drug makes a significant difference.



# Parametric Survival Models

Task: Fit parametric survival models containing treatment (tx) and CD4 count (cd4) as explanatory variables.

Method: Fit exponential, Weibull and log-logistic models, concluded that log-logistic was best has it has the lowest AIC.

| Log-logistic Model Summary | | | | |
|---|---|---|---|---|
| | Value | 95% CI | z | p |
| Intercept | 6.82584 | [6.33, 7.32] | 26.82 | $<$ 2e-16 |
| Treatment | 0.84295 | [0.27, 1.41] | 2.91 | 0.0036 |
| cd4 | 0.02080 | [0.013, 0.028] | 5.55 | 2.9e-08 |

- Time ratio for the treatment effect: The median time to survival time for new treatment is 2.32 times the median for old treatment. 95% CI:[1.32, 4.10]
- Tested Cox-snell residuals and concluded they looked normal