Survival with malignant melanoma

Rasmus Waagepetersen

27. november 2024

The data file melanoma30 contains data on 205 patients operated for malignant melanoma at Odense University Hospital in the period 1962 – 1973 and followed until the end of 1977 (Drzewiecki and Andersen 1982). Thus survival times were censored for patients who were alive at the end of study.

For each patient the following informations are recorded:

1. Patient number `id`
2. Time in study `time`
3. Cause of death `status`
4. Dead/alive at time of follow-up `dead`
5. Inflammatory cell infiltration score `ici`
6. Presence of epitheloid cells `epicell`
7. Presence of ulceration `ulceration`
8. Thickness of tumor (1/100mm) `thickness`
9. Sex `sex`
10. Age at operation `age`
11. Natural log of thickness/100 `logthick`
12. Depth of invasion of tumor score `invas2`

In this analysis we will consider time to death. The primary hypothesis to be investigated is whether tumor thickness has an impact of survival. If so this impact should be quantified. Death can be due to malignant melanoma or other causes (cf. variable `status`) but in this analysis we will not distinguish between different causes of death (in which case we would face a competing risk scenario).

Tasks:

1. (a) What type of censoring takes place for this study when considering survival to death of any cause. Can you assume independent censoring ?

   (b) Suppose death from other causes was also considered a censoring mechanism (i.e. if we only considered survival to death of malignant melanoma). Would independent censoring hold in that case ?

2. Assess each of the variables. Are there issues with outliers, small groups for categorical variables, skew distributions for continuous variables,... ? Next check whether the different variables are pairwise correlated/associated.

3. Investigate to begin with the effect of tumor thickness by converting it to a categorical variable. Consider estimates of the survival functions for the thickness categories and by performing a log-rank test.

4. Fit a Cox proportional model with all covariates (main effects) included. Investigate whether some kind of transformations should be applied to the continuous covariates age and tumor thickness. Here you can use martingale residuals, parameter estimates for categorized covariates, comparison of models obtained with different transformations of the covariates etc. Interpret the parameter estimates for all the covariates in the model.

5. Perform a careful assessment of the fitted model using the whole battery of methods for checking the Cox proportional hazards model.

6. Within the framework of the fitted Cox regression model, possibly with transformations of age and thickness, test the null hypothesis of no effect of tumor thickness.

7. Give an estimate for the survival function for a 57 year old man with `ici` score 2, presence of epitheloid cells and ulceration, thickness of tumor 750, and invasiveness I-III. What is the probabibility that he wil survive to be 65 ?

8. Is there scope for applying a parametric model - e.g. exponential or Weibull ? (consider appropriate plot of estimated cumulative baseline hazard).