Recitation 3

ABA Spring 2024
02/09/24

Agenda

- Cox Regression example

Reminders

- HW1 due Feb. 12, 11:59PM
- Questions?

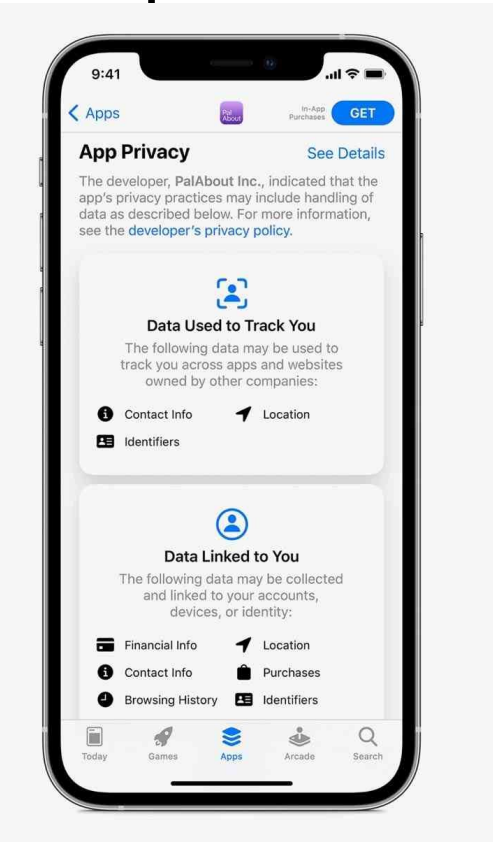
Motivation for Maximum likelihood estimation

- KM provided us with useful information about survival
- We need a “regression” like model to account for covariates
- We’ll use the widely studied Cox Proportional Hazard model (Cox Regression)
- Cox is a semi-parametric approach:

  We estimate the parameters (Betas) of the model using the partial likelihood.

Data for Cox Regression: App Store example

- The Apple app store requires that all apps in the store must display a privacy label. The goal of the privacy label is to inform users about the app's privacy posture and let them make decision about whether to download the app.
- There has been a lot of interest in whether privacy labels affect demand for the app.
- From December 14 onwards, the App Store mandated that any new or any existing app must display a privacy label.



Data

- I am attaching a dataset on the date when the Apps display the label.
- Apps are followed for a certain number of days and if the label appears then it is listed as 1, otherwise 0.
- One would expect that labels will appear on or around Dec 14.
- However, that does not happen. Existing apps do not display privacy labels despite the App Store mandate.
- This leads to additional investigating on what factors cause apps to display labels.

Event of interest: When does the app display the privacy label?

Data

In this question, we examine this issue carefully. While there are many useful attributes that may influence the timing for the labels, we focus on:

- rank of the app on Dec 14 and
- the type of the app (Free, paid, grossing).

The App Store publishes the rank of apps based on how many downloads they receive. A higher rank (1, 2, 3.. and so on) app gets more downloads.

We have reasons to believe that top ranked apps are more likely to disclose the label earlier.

Data

- The data provides information on the number of days it takes an App to display the Label.
- The column “days_followed” lists the number of days the app was followed. The “Label“ column shows whether the label appeared

.

Interpreting the estimates

- The interpretation in Cox models follows regression framework. Recall we are modeling "time to event" for a subject.
- Hazard rate signals propensity of then even to occur (or fail).
- Recall, we have
  - $\log\big(h(t,X)\big) = \log\big(h_0(t)\big) + X\beta$
- Or,
  - $\log\big(h(t,X)\big) - \log\big(h_0(t)\big) = X\beta$
- Or $\log\left(\frac{h(t,X)}{h_0(t)}\right)$=Xβ
  - β indicates the impact of a unit increase in X on log of hazard rate.
- Notice that we do not assume a functional form for hazard. We cannot say how X affects y (time to event). We can only say how X affects hazard relative to the baseline hazard $h_0(t)$.
  - So, the impact of covariate is measured as hazard ratio.

Interpreting the estimates

- A unit change in X causes β unit change in the log of hazard ratio.
- We can exponentiate the coefficients, exp(β) to get the change in the hazard ratio.
- When
  - exp(β) > 1, hazard of time to event has gone up due a unit change in X. If exp(β) = 1.5, then hazard has increased by (1.5 - 1) = 50%.
  - exp(β) = ~~0~~, there is no change in hazard, exp(beta) =1
  - exp(β) < 1, hazard has gone down. If exp(β) = 0.8, then hazard has decreases by (1 -0.8) = 20%.