Chittumuri_Stat755_HW4

Isabella Chittumuri

4/21/2021

```
setwd("~/Documents/Hunter College/Spring 2021/Stat 755/HW")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(survival)
library(survminer)
```

```
## Loading required package: ggpubr
```

The following questions consider a dataset from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," Med. J. Aust., 1991). These data comprise the times in days spent by heroin addicts from entry to departure from one of two methadone clinics. There are two additional covariates, namely, prison record and maximum methadone dose, believed to affect the survival times. The dataset name is addicts.dat. A listing of the variables is given below:

Column 1: Subject ID Column 2: Clinic (1 or 2) Column 3: Survival status (0 1/4 censored, 1 1/4 departed from clinic) Column 4: Survival time in days Column 5: Prison record (0 1/4 none, 1 1/4 any) Column 6: Maximum methadone dose (mg/day)

1. The following edited printout was obtained from fitting a Cox PH model to these data:

```
knitr::include_graphics('resources/1. Cox PH Model.png', dpi = 100)
```

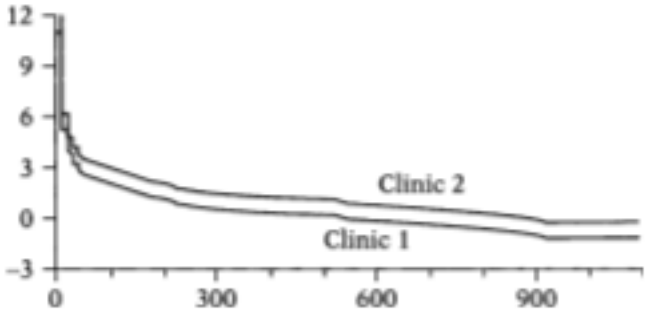
Cox regression Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]		P(PH)
Clinic	−1.009	0.215	0.000	0.365	0.239	0.556	0.001
Prison	0.327	0.167	0.051	1.386	0.999	1.924	0.332
Dose	−0.035	0.006	0.000	0.965	0.953	0.977	0.347
No. of subjects: 238			Log likelihood = −673.403				

Based on the information provided in this printout, what do you conclude about which variables satisfy the PH assumption and which do not? Explain briefly.

The P(PH) value for Prison is 0.332 and Dose is 0.347, which are both greater than the alpha level of 0.05. This means that they both satisfy the PH assumption. However, the P(PH) value for CLinic is 0.001, which is less than the alpha level of 0.05. Therefore, Clinic does not satisfy the PH assumption.

2. Suppose that for the model fit in question 1, log–log survival curves for each clinic adjusted for prison and dose are plotted on the same graph. Assume that these curves are obtained by substituting into the formula for the estimated survival curve the values for each clinic and the overall mean values for the prison and dose variables. Below, we show these two curves. Are they parallel? Explain your answer.

```
knitr::include_graphics('resources/2. log-log curves.png')
```



The log-log plots are parallel, but this is expected because clinic was already included in the model which assumed to satisfy the PH assumption. If clinic was not included in the model, the log-log plots may not be parallel when comparing the two clinics adjusted for prison and dose.

3 The following printout was obtained from fitting a stratified Cox PH model to these data, where the variable being stratified is clinic:

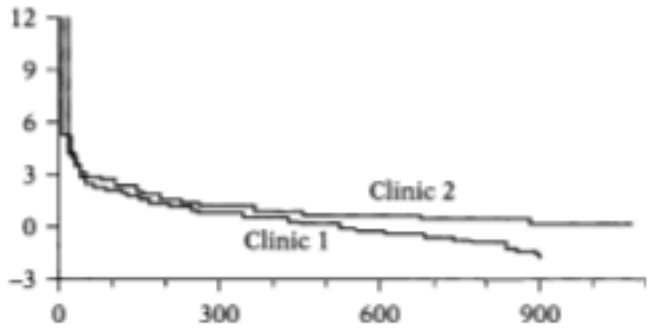
```
knitr::include_graphics('resources/3. Stratified Cox PH.png', dpi = 100)
```

Stratified Cox regression Analysis time.t: survt (in days)	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]
Prison	0.389	0.169	0.021	1.475	1.059	2.054
Dose	−0.035	0.006	0.000	0.965	0.953	0.978

No. of subjects = 238 Log likelihood = −597.714 Stratified by clinic

Using the above fitted model, we can obtain the log–log curves below that compare the log–log survival for each clinic (i.e., stratified by clinic) adjusted for the variables prison and dose. Using these curves, what do you conclude about whether or not the clinic variable satisfies the PH assumption? Explain briefly.

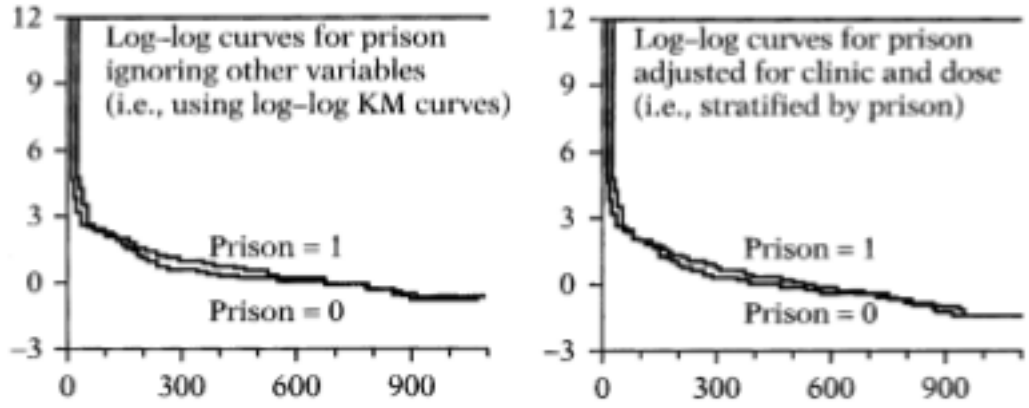
```
knitr::include_graphics('resources/3. log-log curves.png')
```



When clinic is stratified, the log-log plots are not parallel. They intersect early on and then diverge towards the end.

4. Consider the two plots of log–log curves below that compare the log–log survival for the prison variable ignoring other variables and adjusted for the clinic and dose variables. Using these curves, what do you conclude about whether or not the prison variable satisfies the PH assumption? Explain briefly.

```
knitr::include_graphics('resources/4. log-log suvival.png')
```



Looking at the log-log curves, we see that for both graphs the curves intersect, diverge from each other, and merged into one line multiple times. Therefore, they are not parallel and do not satisfy the PH assumption.

5. How do your conclusions from question 1 compare with your conclusions from question 4? If the conclusions differ, which conclusion do you prefer? Explain.

My conclusion in question 1 was that prison and dose satisfy the PH assumption based on the P(PH) value. But my conclusion in question 4 was that prison does not satisfy the PH assumption based on the log-log curves. I prefer second conclusion where I can visually see how the variable plots in the log-log curves. And I saw that even though the P(PH) value was high, I still saw that the log-log curves intersect many times. This solidifies that the prison does not satisfy the PH assumption.

6. Describe briefly how you would evaluate the PH assumption for the variable maximum methadone dose using observed versus expected plots.

Since maximum methadone dose is a continuous variable, we have to categorize it into groups. For the observed plot, we can divide the variable into two groups for high and low and then plot the KM curves for each group. For the expected plot, we can use the Cox PH model containing the dose variable, and substitute the mean or median dose for each group into the formula for the estimated survival curve for each group. Lastly, we would compare the observed and expected plots to see if there are parallel

7. State an extended Cox model that would allow you to assess the PH assumption for the variables clinic, prison, and dose simultaneously. For this model, state the null hypothesis for the test of the PH assumption and describe how the likelihood ratio statistic would be obtained and what its degrees of freedom would be under the null hypothesis.

Extended Cox model for clinic, prison and dose:

$$h(t, X) = h_0(t)e^{\beta_1 Clinic + \beta_2 Prison + \beta_3 Dose + \delta_1(Clinic * g(t)) + \delta_2(Prison * g(t)) + \delta_3(Dose * g(t))}$$

Null Hypothesis $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$

Likelihood Ratio Statistic $LR = -2lnL_R - (-2lnL_F)$

L_F is the full model L_R is the reduced model, where $\delta_1 = \delta_2 = \delta_3 = 0$

The LR statistic is approximately χ^2 under the H_0 with 3 degrees of freedom.

8. State at least one drawback to the use of the extended Cox model approach described in question 7.

One drawback to using the extended Cox model approach is that it isn't clear how to specify $g(t)$. Different choices can lead to different conclusions.

9. State an extended Cox model that would allow you to assess the PH assumption for the variable clinic alone, assuming that the prison and dose variables already satisfy the PH assumption. For this model, state the null hypothesis for the test of the PH assumption, and describe how the likelihood ratio (LR) statistic would be obtained. What is the degrees of freedom of the LR test under the null hypothesis?

Extended Cox model for clinic, when prison and dose satisfy PH assumption

$$h(t, X) = h_0(t)e^{\beta_1 Clinic + \beta_2 Prison + \beta_3 Dose + \delta_1(Clinic * g(t))}$$

Null Hypothesis $H_0 : \delta_1 = 0$

Likelihood Ratio Statistic $LR = -2lnL_R - (-2lnL_F)$

The LR statistic is approximately χ^2 under the H_0 with 1 degrees of freedom.

10. Consider the situation described in question 9, where you wish to use an extended Cox model that would allow you to assess the PH assumption for the variable clinic alone, assuming that the assumption is satisfied for the prison and dose variables. Suppose you use the following extended Cox model:

$$h(t,X) = h_0(t)e^{\beta_1 Prison + \beta_2 Dose + \beta_3 Clinic + \delta_1 (Clinic) g(t))}$$

$$Where: g(t) = 1 if t > 365 days$$
$$g(t) = 0 if t \leq 365 days$$

For the above model, what is the formula for the hazard ratio that compares clinic 1 to clinic 2 when t is greater than 365 days? when t is less than or equal to 365 days? In terms of the hazard ratio formula just described, what specific departure from the PH assumption is being tested when the null hypothesis is $H_0 : \delta_1 = 0$?

$$HR = \frac{h(t, Clinic = 2)}{h(t, Clinic = 1)} = \frac{h_0(t)e^{\beta_1 Clinic + \beta_2 Prison + \beta_3 Dose + \delta_1 (Clinic * g(t))}}{h_0(t)e^{\beta_1 Clinic + \beta_2 Prison + \beta_3 Dose + \delta_1 (Clinic * g(t))}}$$
$$= e^{\beta_1 + \delta_1 g(t)}$$

Note: Prison and Dose cancel out because we are only controlling for Clinic

$$For t > 365 days : g(t) = 1$$
$$HR = e^{\beta_1 + \delta_1}$$
$$For t \leq 365 days : g(t) = 0$$
$$HR = e^{\beta_1}$$

If $\delta_1 \neq 0$, then model assumes the hazard ratio is not constant over time and that it does not satisfy the PH assumption for the clinic variable.