Chittumuri_Stat755_HW5

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

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. Two other covariates, namely, prison record and maximum methadone dose, are 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 P(PH) information in the above printout, it appears that clinic does not satisfy the PH assumption; this conclusion is also supported by comparing log–log curves for the two clinics and noticing strong non-parallelism. What might we learn from fitting a stratified Cox (SC) model stratifying on the clinic variable? What is a drawback to using a SC procedure that stratifies on the clinic variable?

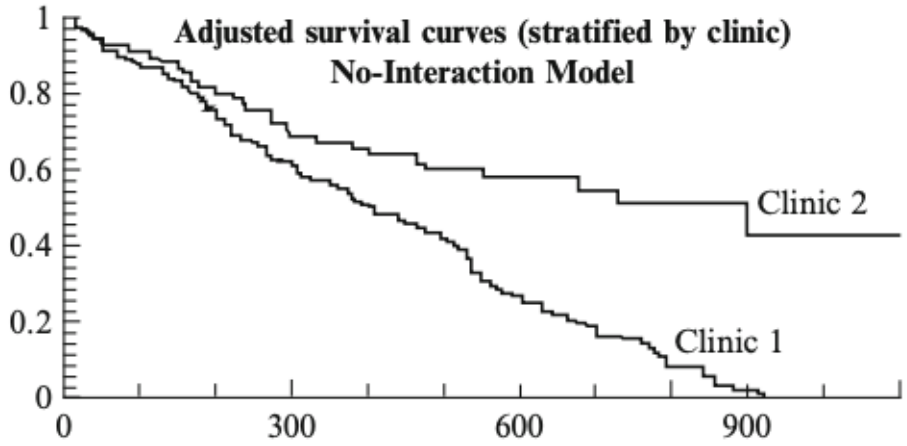
- If we fit a stratified Cox model stratifying on the clinic, we can compare adjusted survival curves for both clinic 1 and 2, whiling adjusting for prison and dose. This lets us visualize the difference the clinics survival curves. One drawback to using an SC procedure is that we cannot estimate the hazard ratio for clinic since it's included in the model.

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

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 adjusted curves below that compare the adjusted survival probabilities for each clinic (i.e., stratified by clinic) adjusted for the variables, prison and maximum methadone dose.



Based on these adjusted survival curves, what conclusions can you draw about whether the survival experience is different between the two clinics? Explain.

- The adjusted survival curve for clinic 2 is consistently higher than that of clinic 2. The difference between the two grow over time. This plot suggested that over time clinic 2 has a better survival outcome than clinic 1.

3. State the hazard function model being estimated in the above computer results. Why is this model a no-interaction model?

$$h(t, X) = h_0(t)e^{\beta_1 Prison + \beta_2 Dose}, \quad g = 1, 2$$

- This is a no-interaction model because the coefficients for the variables prison and dose are the same for each stratum

4. Using the above computer results, provide point and interval estimates for the effect of prison adjusted for clinic and dose. Is this adjusted prison effect significant? Explain.

The effect of prison adjusted for clinic and dose is

$$\hat{HR} = 1.475$$
$$CI = (1.059, 2.054)$$

- Yes, this adjusted prison effect is significant because it has a p-value of 0.021, which is less than alpha level 0.05.

5. The following computer results consider a SC model that allows for interaction of the stratified variable clinic with each of the predictors, prison and dose. Product terms in the model are denoted as clinpr = clinic x prison and clindos = clinic x dose.

Stratified Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	P > |z|	Haz. Ratio	[95% Conf.	Interval]
prison	1.087	0.539	0.044	2.966	1.032	8.523
dose	−0.035	0.020	0.079	0.966	0.929	1.004
clinpr	−0.585	0.428	0.172	0.557	0.241	1.290
clindos	−0.001	0.015	0.942	0.999	0.971	1.028

No. of subjects = 238 Log likelihood = −596.779 Stratified by clinic

State two alternative versions of the interaction model being estimated by the above printout, where one of these versions should involve the product terms used in the above printout.

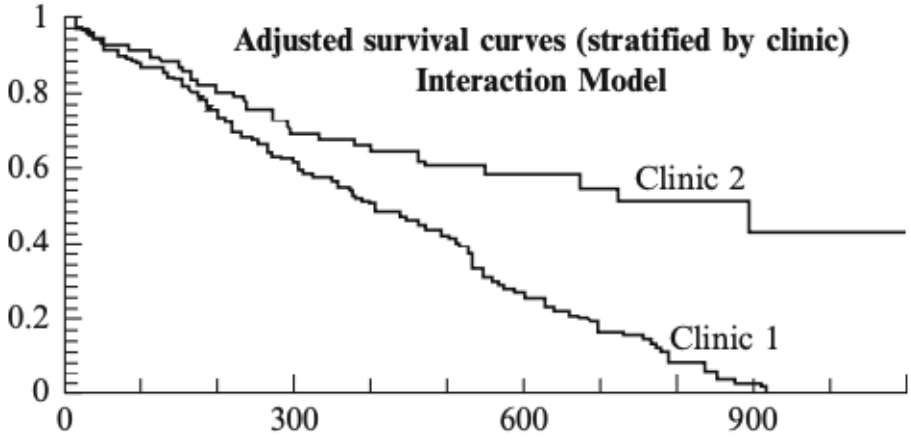
$$Version1 : h(t, X) = h_{0g}(t)e^{\beta_{1g}Prison + \beta_{2g}Dose}, \quad g = 1, 2$$

$$Version2 : h_g(t, X) = h_{0g}(t)e^{\beta_1 Prison + \beta_2 Dose + \beta_3(CLinic * Prison) + \beta_4(Clinic * Dose)}, \quad g = 1, 2$$

6. Using the computer results above, determine the estimated hazard models for each clinic. (Note that the clinics are coded as 1 or 2.)

$$g = 1(clinic1) :$$
$$h_1(t, X) = h_{01}(t)e^{(0.502)Prison + (-0.036)Dose}$$
$$g = 2(clinic2) :$$
$$h_2(t, X) = h_{02}(t)e^{(-0.083)Prison + (-0.037)Dose}$$

7. Below are the adjusted survival curves for each clinic based on the interaction model results above. These curves are adjusted for the prison and dose variables.



Compare the survival curves by clinic obtained for the interaction model with the corresponding curves previously shown for the no-interaction model. Do both graphs indicate the similar conclusions about the clinic effect? Explain.

- The survival curves by clinic for both the interaction model and no-interaction model are visually the same. Therefore, both graphs indicate the same conclusion that overtime clinic 2 has a consistently and increasingly higher survival curve than that of clinic 1. This means that clinic 2 has a higher probability for survival than clinic 1.

8. Carry out a likelihood ratio test to determine whether the no-interaction model is appropriate. In doing so, make use of the computer information described above, state the null hypothesis, state the form of the likelihood statistic and its distribution under the null hypothesis, and compute the value of the likelihood statistic and evaluate its significance. What are your conclusions?

$H_0 : \beta_3 = \beta_4 = 0$

$H_a : \beta_3 \neq \beta_4 \neq 0$

LR: $= -2lnL_R - (-2lnL_F) = (-2 * -597.714) - (-2 * -596.779) = 1195.428 - 1193.558 = 1.18$

R refers to no-interaction model and F refers to the interaction model

LR of 1.87 is approximately χ^2 with 2 d.f. under H_0, which is not significant. Therefore, the no-interaction model is the better model.