Lin_ST625_HW8

Frances Lin

2/25/2022

1

ID	t_i	delta	Z_i
1	6	1	3.28
1	7	0	3.2
1	9	1	2.8
1	10	1	2.7
1	12	0	2.64
1	15	1	2.05

1a

The formula for the Cox PH (proportional hazard) model is given as $h(t|Z) = h_0(t)exp(Z^T\beta)$, where $h(t|Z)$ is the hazard function at time t given a set of covariates, $h_0(t)$ is the baseline hazard function when all covariates are at their baseline values.

Or equivalently, $S(t|Z) = S_0(t)exp(Z^T\beta)$.

1b

The partial likelihood is given as $\phi_1(\beta)$ $\phi_2(\beta)$ $\phi_3(\beta)$ $\phi_4(\beta)$, where

$\phi_1(\beta) = \frac{e^{(3.28\beta)}}{e^{(3.28\beta)}+e^{(3.2\beta)}+e^{(2.8\beta)}+e^{(2.7\beta)}+e^{(2.64\beta)}+e^{(2.05\beta)}}$,

$\phi_2(\beta) = \frac{e^{(2.8\beta)}}{e^{(2.8\beta)}+e^{(2.7\beta)}+e^{(2.64\beta)}+e^{(2.05\beta)}}$,

$\phi_3(\beta) = \frac{e^{(2.7\beta)}}{e^{(2.7\beta)}+e^{(2.64\beta)}+e^{(2.05\beta)}}$,

$\phi_4(\beta) = \frac{e^{(2.05\beta)}}{e^{(2.05\beta)}}$.

t_i	risk_set	Z_i
1	1*,2,3,4,5,6	3.28
3	3*,4,5,6	2.8
4	4*,5,6	2.7
6	6*	2.05

1. Order the event time (e.g. 6, 9, 10, 15)
2. Create a risk set and highlight the individual

3. Identify corresponding Z_i
4. Get ϕ_i and multiply them

2a

No, one of the covariates (`Size1`, `Size2` or `Size3`) will serve as the baseline or reference group.

2b

The formula for the Cox PH (proportional hazard) model is given as $h(t|Z) = h_0(t)exp(Z^T\beta) = h_0(t)exp(\beta_1 P27 + \beta_1 CYCLINE + \beta_2 NODES + \beta_3 SIZE2 + \beta_4 SIZE4 + \beta_5 Age + \beta_6 Year)$.

2c

The hazard have to be proportional for the relative hazard to remain constant over time for different covariates or covariate levels.

2d

The survival times depend significantly on all variables but Age and Year at the $\alpha = 0.05$ level.

2e

No, patients with abnormal CYCLINE do not have longer expected survival time. The estimate for `CYCLINE` indicates that the risk of death is higher if the patient falls in the case of `CYCLINE` $= 1$.

2f

No, patients with larger tumor size do not have longer expected survival time. This can be seen from the estimate for `SIZE2` and `SIZE3`. Both values indicate that the risk of death is higher for both cases. (The hazard ratio of `SIZE2` vs `SIZE1` > 1. The hazard ratio of `SIZE3` vs `SIZE1` also > 1.)

2g

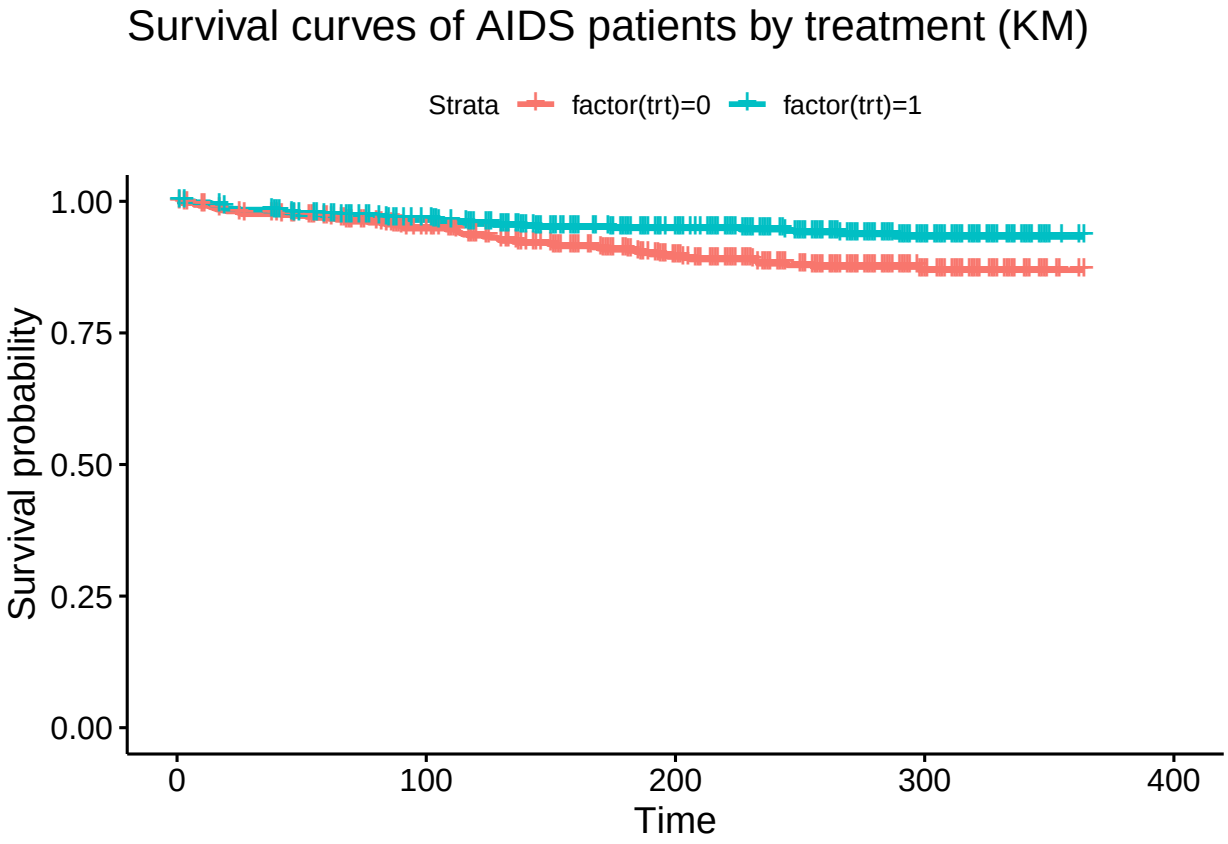
It is estimated that the risk of death is $exp(0.7256) = 2.06597$ times higher for the other group (protein CYCLINE is abnormal: `CYCLINE` $= 1$), as compared to the reference group (`CYCLINE` $= 0$).

2h

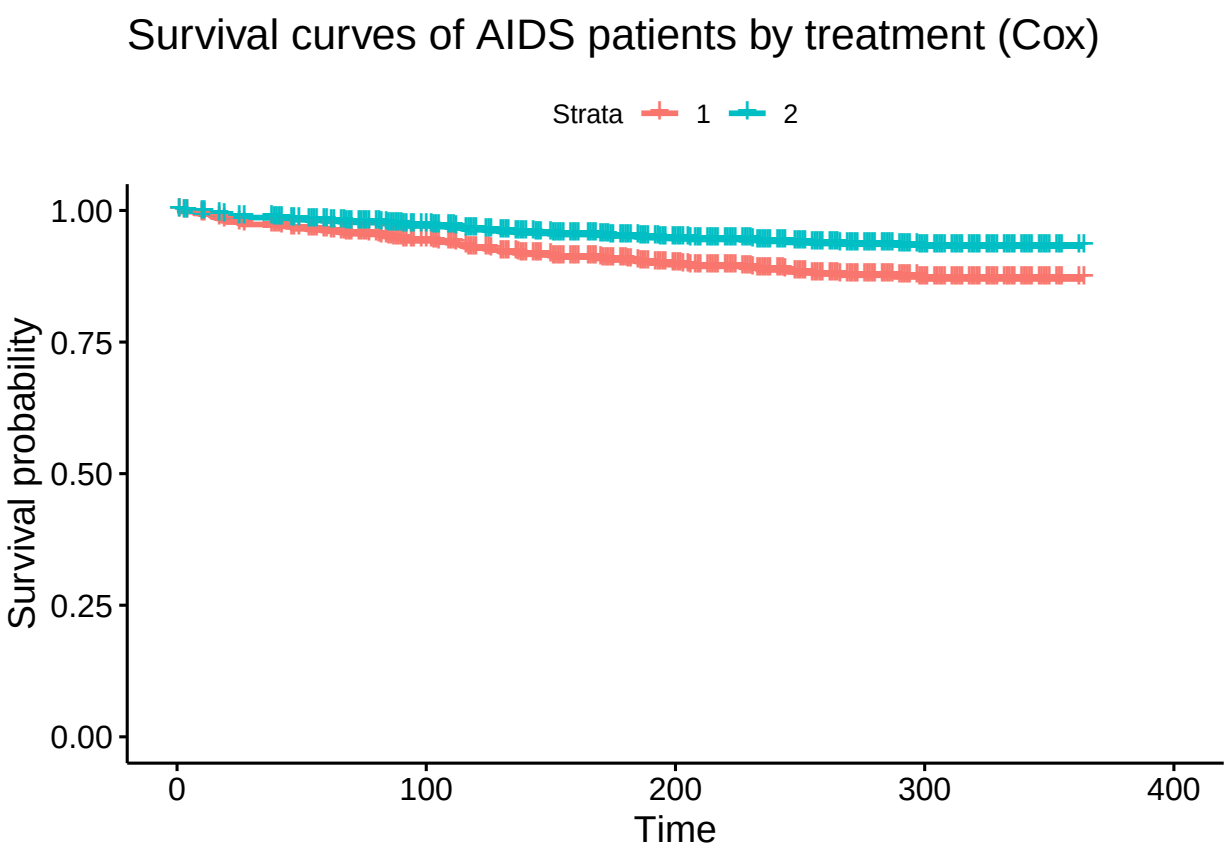
It is estimated that the risk of death is $exp(0.6680) = 1.950333$ times higher for the other group (tumor size is between 2cm and 4cm: `SIZE2` $= 1$), as compared to the reference group (`SIZE2` $= 0$).

3

3a


Survival curves of AIDS patients by treatment (KM)
Strata
factor(trt)=0
factor(trt)=1
Survival probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Time

3b



3c

The two survival plots differ but they don't differ much. KM estimator is a nonparametric method. Cox PH model is a semiparametric regression model so it does assume more assumptions as compared to the KM estimator, but it is more flexible than the AFT model. Specifically, Cox PH model is nonparametric for $h_0(t)$ and parametric for $exp(Z^T\beta)$.

3d

Cox PH model results with `race` (1 if white, 2 if black, 3 if Hispanic, 4 if Asian, 5 if native American) only and treat `race` as a categorical variable:

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
factor(race)2	-0.2257	0.7979	0.2594	-0.8702	0.3842
factor(race)3	0.1901	1.209	0.2639	0.7206	0.4712
factor(race)4	1.092	2.98	0.5942	1.837	0.06615
factor(race)5	0.05464	1.056	1.01	0.05411	0.9569

Since the formula is given as $h(t|Z) = h_0(t)exp(Z^T\beta) = h_0(t)exp(\beta_1 factor(race)2 + \beta_2 factor(race)3 + \beta_3 factor(race)4 + \beta_4 factor(race)5)$, the MPLE (maximum partial likelihood estimator) of the hazard ratio comparing "blacks" (`race` $= 2$) to "Hispanics" (`race` $= 3$) is $\frac{e^{(Z0.1901)}}{e^{(Z0.05464)}+e^{(Z(-0.2257))}}\ \frac{e^{(Z(-0.2257))}}{e^{(Z(-0.2257))}}$. I am not confident about this one.

It is estimated that the risk of death is $exp(-0.22575)/exp(0.19012) = 0.7979/1.2094 = 0.6597$ times lower for black (`race` = 2) patients, as compared to Hispanics (`race` = 3) patients.

This makes sense since $\frac{h(t)_{race=2}}{h(t)_{race=1}} = exp(-0.22575) = 0.7979$ (i.e. Black has lower risk than white.) On the other hand, $\frac{h(t)_{race=3}}{h(t)_{race=1}} = exp(0.19012) = 1.2094$ (i.e. Hispanics has higher risk than white.)

3e

Treating `trt`, `sex`, `IV`, `hamophiliac` and `KPS` as categorical variables, the results shows that `trt`, `CD4_count` and `KPS` have significant effects on survival at the $\alpha = 0.05$ level.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
factor(trt)1	-0.6738	0.5098	0.216	-3.12	0.001809
CD4_count	-0.0146	0.9855	0.002564	-5.693	1.248e-08
factor(sex)2	0.131	1.14	0.2853	0.4593	0.646
factor(IV)2	0.7313	2.078	1.027	0.7121	0.4764
factor(IV)3	-0.622	0.5369	0.3364	-1.849	0.06444
factor(hamophiliac)1	0.1096	1.116	0.601	0.1824	0.8552
factor(KPS)80	-0.4452	0.6407	0.3688	-1.207	0.2274
factor(KPS)90	-1.131	0.3228	0.3691	-3.063	0.002191
factor(KPS)100	-1.566	0.2088	0.4126	-3.796	0.0001471
ZDV	-0.0004452	0.9996	0.003856	-0.1155	0.9081
age	0.02192	1.022	0.01138	1.927	0.05404

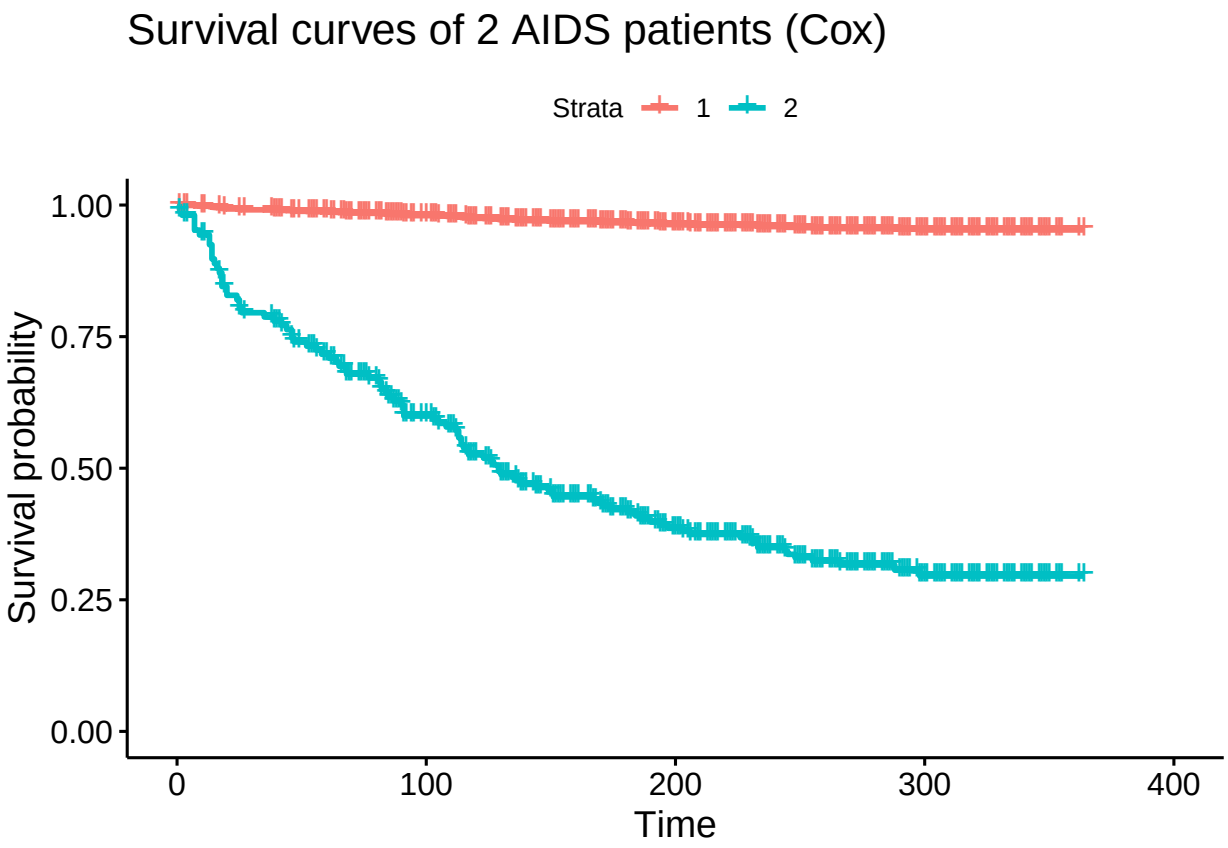
It is estimated that the risk of death is $exp(-0.6737878) = 0.5097$ times lower for the other group (IDV: `trt` = 1), as compared to the reference group (control: `trt` = 0).

It is estimated that the risk of death decreases by $(1 - exp(-0.0145956)) * 100 = 1.4489\%$ with 1-unit increase in the CD4 count in cells.

It is estimated that the risk of death is $exp(-1.1305838) = 0.3228$ times lower for the other group (minor symptoms: `KPS` = 90), as compared to the reference group (active work impossible: `KPS` = 70).

It is estimated that the risk of death is $exp(-1.5662040) = 0.2088$ times lower for the other group (no evidence of disease: `KPS` = 100), as compared to the reference group (active work impossible: `KPS` = 70).

3f



Patients information is as follows

id	trt	CD4_count	sex	IV	hamophiliac	KPS	ZDV	age
1	1	86	2	1	0	90	30	38
2	0	20	2	2	1	70	250	25