

Also see Quiz 10, 2021

Math 473 Final Spring 2017

Name _____

variables	AIC
treat age serum size gleas	32.17
treat age size gleas	30.17
treat size gleas	28.57 ← -7
size gleas	27.53
size	31.04

	coef	exp(coef)	se(coef)	z	p
treat	-1.1127	0.3287	1.2031	-0.92	0.355
size	0.0826	1.0861	0.0475	1.74	0.082
gleas	0.7102	2.0345	0.3379	2.10	0.036

Likelihood ratio test=13.8 on 3 df, p=0.00323
n= 38, number of events= 6

1) Data is from Collett (2003, p. 10) regarding survival of prostate cancer patients. The predictors are *treat* (1 if placebo, 2 if DES), *age* = patient age, *serum* (a prognostic variable), *size* of the primary tumour, and *gleas* (Gleason index: the more advanced the tumour, the higher the index). Results from backward elimination are shown.

a) What is the best starting submodel I_{II} ?

$$I_{II} = I_{min} = \boxed{\text{size, gleas}}$$

b) Are there any other candidate submodels? Explain briefly.

$$AIC(\text{size}) = 31.04 < AIC(I_{min}) + 7 = 34.53$$

$$\boxed{\text{size}}$$

c) Consider the model containing $x_1 = \text{treat}$, $x_2 = \text{size}$ and $x_3 = \text{gleas}$. Find the $ESP = \hat{\beta}^T x$ if $x_1 = 2$, $x_2 = 20$, and $x_3 = 10$.

$$= -1.1127(2) + 0.0826(20) + 0.7102(10)$$

$$= \boxed{6.5296}$$

	M1	M2	M3	M4
# of predictors	10	3	2	1
# with $0.01 \leq p\text{-value} \leq 0.05$	2	2	1	1
# with $p\text{-value} > 0.05$	8	1	0	0
$-2\log(L)$	419.470	422.708	425.704	429.795
$AIC(I)$	439.470	428.708	429.704	431.795
p-value for change in PLR test	1.0	0.862	0.304	0.325

06d21 2) The above table gives summary statistics for 4 PH regression models considered as final submodels after performing variable selection. Assume that the PH assumptions hold for all 4 models. The full model was M1, and M2 was the minimum AIC model found. Which model should be considered as the first starting submodel $I_{\mathcal{J}}$? Explain briefly why each of the other 3 submodels should not be used as the starting submodel.

$$429.704 = AIC(M3) < AIC(I_{\min}) + 2 = 430.708$$

$$AIC(M4) > 430.708 \quad \text{SO}$$

$$I_{\mathcal{J}} = M3$$

M1 and M2 have too many predictors

M4 has an AIC that is too large

OR $AIC(M4) > AIC(I_{\min}) + 2$

↑
→ -2

15

3) Suppose

$$S(t) = 1 - \left(\frac{t}{\omega}\right)^\theta$$

Eldzi

for $0 < t < \omega$ where $\theta > 0$. Find the following quantities for $0 < t < \omega$.

$$\text{a) } F(t) = 1 - S(t) = \left(\frac{t}{\omega}\right)^\theta$$

$$\text{b) } f(t) = F'(t) = -S'(t) = \frac{\theta}{\omega} \left(\frac{t}{\omega}\right)^{\theta-1} = \frac{\theta t^{\theta-1}}{\omega^\theta}$$

$$\text{c) } h(t) = \frac{f(t)}{S(t)} = \frac{\frac{\theta}{\omega} \left(\frac{t}{\omega}\right)^{\theta-1}}{1 - \left(\frac{t}{\omega}\right)^\theta} = \frac{\theta t^{\theta-1}}{\omega^\theta - t^\theta}$$

$$\text{d) } H(t) = -\log S(t)$$

$$= -\log \left[1 - \left(\frac{t}{\omega}\right)^\theta \right]$$

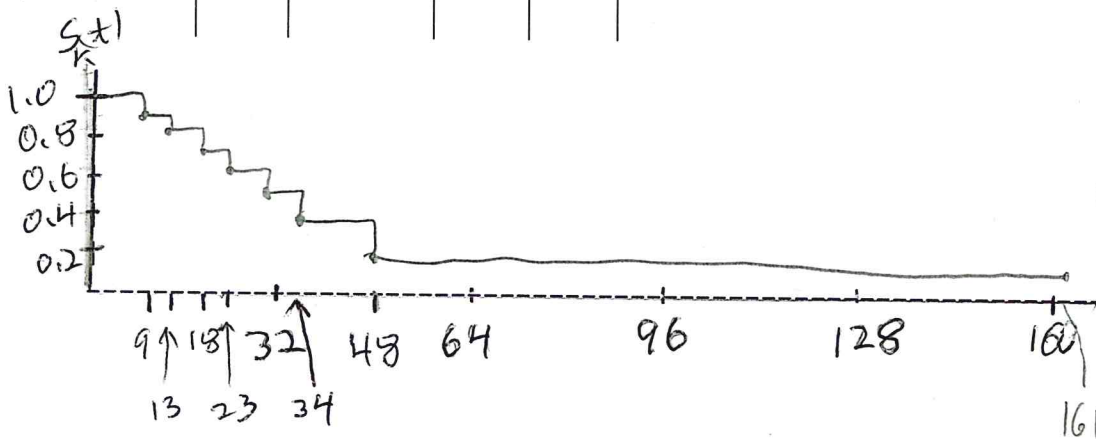
Eid 21

4) Survival times for 11 patients with AM leukemia are given below. The data is from Smith (2002, p. 115) and Miller (1981, p. 49).

9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+

Compute the Kaplan Meier survival function $\hat{S}_K(t_i)$ by filling in the table below. Show what you multiply to find $\hat{S}_K(t_i)$. Then plot the function.

$t_{(j)}$	γ_j	t_i	n_i	d_i	$\hat{S}_K(t_i) = \hat{S}_K(t_{i-1}) \left(1 - \frac{d_i}{n_i}\right)$
		$t_0 = 0$			$\hat{S}_K(0) = 1$
9	1	9	11	1	$\hat{S}_K(9) = 1 \left(1 - \frac{1}{11}\right) = \frac{10}{11} = 0.9091$
13	1	13	10	1	$\hat{S}_K(13) = 0.9091 \left(1 - \frac{1}{10}\right) = 0.8182$
13	0				
18	1	18	8	1	$\hat{S}_K(18) = 0.8182 \left(1 - \frac{1}{8}\right) = 0.7160$
23	1	23	7	1	$\hat{S}_K(23) = 0.7160 \left(1 - \frac{1}{7}\right) = 0.6137$
28	0				
31	1	31	5	1	$\hat{S}_K(31) = 0.6137 \left(1 - \frac{1}{5}\right) = 0.4910$
34	1	34	4	1	$\hat{S}_K(34) = 0.4910 \left(1 - \frac{1}{4}\right) = 0.3682$
45	0				
48	1	48	2	1	$\hat{S}_K(48) = 0.3682 \left(1 - \frac{1}{2}\right) = 0.1841$
161	0				



undefined ratio for $t > 161$

$\frac{7}{160} = \frac{4}{160} + \frac{3}{160}$

30

	Without	With	Reduced Model
Criterion	Covariates	Covariates	
-2 LOG L	372.483	362.819	
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.6645	3	0.0216

Parameter		Standard			
Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
treat	1	-0.51757	0.31576	2.6868	0.1012
init	1	0.23605	0.07607	9.6287	0.0019
size	1	0.06790	0.10125	0.4498	0.5024

	Without	With	Full Model
Criterion	Covariates	Covariates	
-2 LOG L	372.483	361.946	
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.5372	6	0.1038

$$\chi^2(RIF) = 362.819 - 361.946 = 0.873$$

Parameter		Standard			
Variable	DF	Estimate	Error	Chi-Square	Pr > ChiSq
treat	1	-0.05019	0.62245	0.0065	0.9357
init	1	0.24563	0.14600	2.8304	0.0925
size	1	0.08920	0.20394	0.1913	0.6618
treatlt	1	-0.26254	0.30704	0.7311	0.3925
initlt	1	-0.00742	0.07958	0.0087	0.9257
sizelt	1	-0.00973	0.10392	0.0088	0.9254

E322

e 5) The bladder cancer data is from Collett (2003, p. 364). The PH model is the reduced model and has predictors *treat* (placebo = 1, thiotepa = 2), *init* (initial number of tumours) and *size* (diameter of largest initial tumour in cm). The GCR model is the full model and adds *treatlt* = *treat**log(time), *initlt* = *init**log(time) and *sizelt* = *size**log(time) interactions to test whether the PH assumption is reasonable. A death time of 0 was changed to 0.001.

a) Test whether the reduced model is good. *H0 reduced model is good HA use full model*

$$\chi^2(RIF) = 10.5372 - 9.6645 = 0.8727$$

$$pval = P(\chi^2_3 > 0.873) > 0.25$$

df	0.25
3	4.11

fail to reject *H0*, the reduced model is good

b) Is the PH assumption reasonable?

yes

	coef	exp(coef)	se(coef)	z	p
X1	-0.0148	0.9853	0.0470	-0.32	0.7526
X2	0.0516	1.0530	0.0394	1.31	0.1905
factor(X3)2	0.2197	1.2457	0.7489	0.29	0.7692
factor(X4)3	-2.4437	0.0868	0.9606	-2.54	0.0110
X5	-0.2436	0.7838	0.8859	-0.27	0.7833
X6	-0.0579	0.9437	0.1612	-0.36	0.7193
X7	2.6127	13.6357	0.9176	2.85	0.0044

Likelihood ratio test=22.3 on 7 df, p=0.00223
n= 37, number of events= 17

E2d21

e 6) The above output is for the Collett (2003, p. 367) data for leukemia patients who received a bone marrow transplant. Y is the survival time in days, and the predictors are $x_1 = \text{page} = \text{age of patient}$, $x_2 = \text{dage} = \text{age of donor}$, factor type of leukemia where the three types were coded as indicator variables x_3 and x_4 , $x_5 = \text{preg} = \text{donor pregnancy}$ (0 for no, 1 for yes), $x_6 = \text{index}$, and $x_7 = \text{Gvhd} = \text{graft-versus host-disease}$ (0=no,1=yes) which can cause transplanted cells to attack host cells (potentially fatally). A PH model was used.

a) Test $\beta = 0$. $H_0: \beta = 0$ $H_A: \beta \neq 0$

$$\chi^2(\text{NIF}) = 22.3$$

$$p\text{val} = 0.00223$$

reject H_0 there is a PH survival relationship between Y and the predictors (Page ... Gvhd).

b) Find a 95% CI for β_7 . $\hat{\beta}_7 \pm 1.96 \text{SE}(\hat{\beta}_7) =$

$$2.6127 \pm 1.96(0.9176) = 2.6127 \pm 1.7985$$

$$= [0.8142, 4.4112]$$

c) Do a 4 step test for $H_0: \beta_7 = 0$. $H_A: \beta_7 \neq 0$

$$z_{07} = 2.85$$

$$p\text{val} = 0.00441$$

reject H_0 , Gvhd is needed in the PH survival model given the other

45

6
predictors (Page ... , index) are in the model

	Value	Std. Error	z	p
(Intercept)	15.1449	16.0795	0.942	3.46e-01
age	-0.1291	0.2186	-0.590	5.55e-01
quant	-0.0455	0.0583	-0.782	4.34e-01
Log(scale)	1.7179	0.3103	5.536	3.10e-08

handway

$$\chi^2 = [-2(29.5)] - [-2(-28.9)]$$

$$= 1.2$$

Scale= 5.57 n =20

Loglik(model)= -28.9 Loglik(intercept only)= -29.5

Chisq= 1.1 on 2 degrees of freedom, p= 0.58

7) The R data set Tobin Data uses a lognormal AFT. (Handled like a WPH or Weibull AFT except use "lognormal AFT" instead of WPH in the appropriate conclusion.) The predictors are age, and quant.

ESP

a) Test $\beta = 0$. $H_0: \beta = 0$ $H_A: \beta \neq 0$

$$\chi^2(2) = 1.1$$

$$pval = 0.58$$

fail to reject H_0 there is not a lognormal AFT survival relationship between γ and the predictor age and quant.

b) Test $\beta_2 = 0$. $H_0: \beta_2 = 0$ $H_A: \beta_2 \neq 0$

$$z_{02} = -0.782$$

$$pval = 0.434 = 4.34e-01$$

fail to reject H_0 , quant is not needed in the lognormal AFT survival model given age is in the model

c) Find the ESP = $\hat{\beta}^T \mathbf{x}$ if $x_1 = age = 50$ and $x_2 = quant = 270$.

$$= -0.1291(50) + -0.0455(270)$$

$$= \boxed{-18.74}$$

8) Does the `cox.zph` function output below suggest that the proportional hazards assumption is reasonable?

```
out <- coxph(Surv(X2,X3)~X4+X5+X6,data=blad)
cox.zph(out)
```

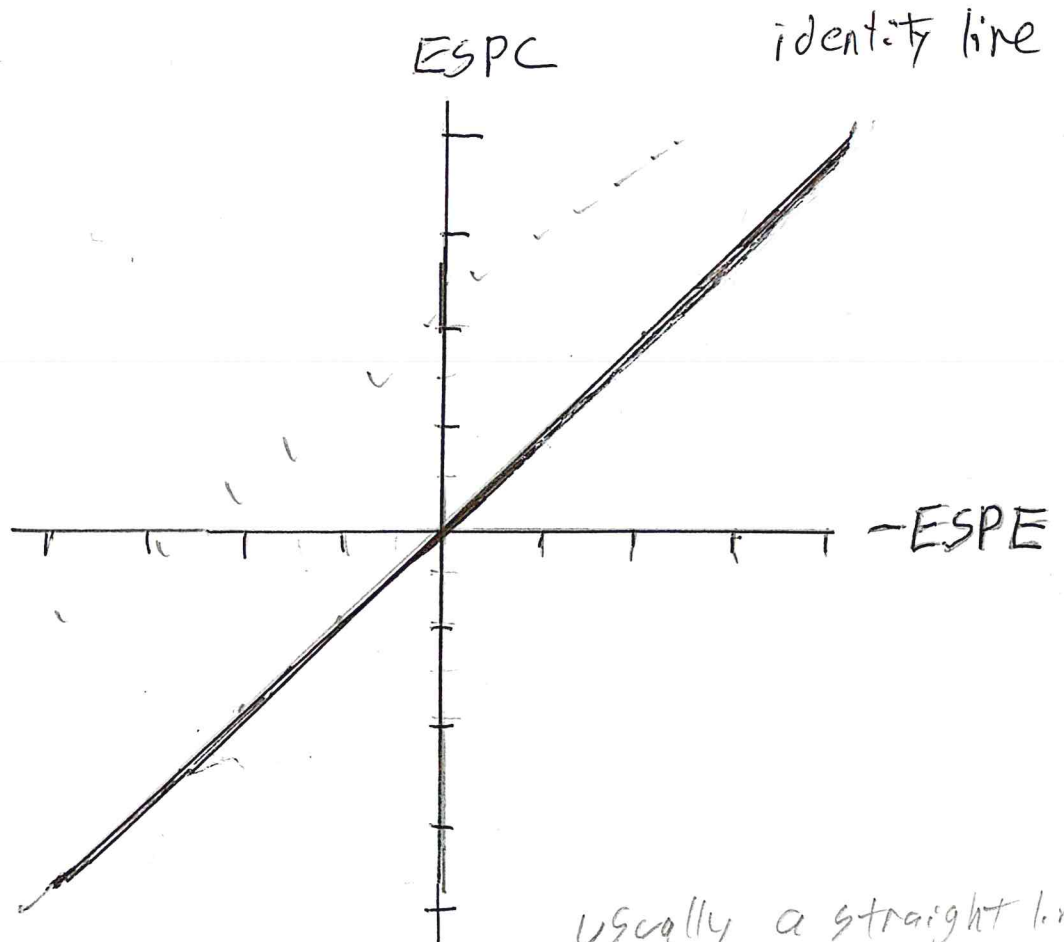
	rho	chisq	p
X4	-0.1234	0.72870	0.393
X5	0.0136	0.00824	0.928
X6	-0.0278	0.04551	0.831
GLOBAL	NA	0.81479	0.846

> 0.05 so

Yes

$p_{\text{val}} < \alpha = 0.05 \Rightarrow$ PH assumption is unreasonable
 $p_{\text{val}} > \alpha = 0.05 \Rightarrow$ PH assumption is reasonable

9) Sketch the EE plot for an exponential PH regression if this model is bad and a Weibull PH model should be used instead.



usually a straight line
not near the identity line

-10 if points tightly cluster about the identity line

15

15