

STA 312 f2023 Quiz 8

1. (4 points) Let $T = e^{\mathbf{x}^T \boldsymbol{\beta}} \times \epsilon$, where $\mathbf{x}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}$, and $\epsilon \sim \exp(1)$. The parameters $\beta_0, \dots, \beta_{p-1}$ are unobserved constants, while the explanatory variable values x_1, \dots, x_{p-1} are observed constants.

Derive the density of T . Show your work. Be sure to indicate where the density is non-zero, and where it is zero.

For $t > 0$ ($\sigma \geq 0$)

$$\begin{aligned}
 f_T(t) &= \frac{d}{dt} F_T(t) = \frac{d}{dt} P(T \leq t) = \frac{d}{dt} P(e^{\mathbf{x}^T \boldsymbol{\beta}} \epsilon \leq t) \\
 &= \frac{d}{dt} P(\epsilon \leq t e^{-\mathbf{x}^T \boldsymbol{\beta}}) = \frac{d}{dt} F_\epsilon(t e^{-\mathbf{x}^T \boldsymbol{\beta}}) \\
 &= f_\epsilon(t e^{-\mathbf{x}^T \boldsymbol{\beta}}) e^{-\mathbf{x}^T \boldsymbol{\beta}} \\
 &= e^{-t e^{-\mathbf{x}^T \boldsymbol{\beta}}} I(t e^{-\mathbf{x}^T \boldsymbol{\beta}} \geq 0) e^{-\mathbf{x}^T \boldsymbol{\beta}} \\
 &= e^{-\mathbf{x}^T \boldsymbol{\beta}} e^{-e^{-\mathbf{x}^T \boldsymbol{\beta}} t} I(t \geq 0), \text{ or } \leftarrow \text{this is 0 if }
 \end{aligned}$$

$$f_T(t) = \begin{cases} e^{-\mathbf{x}^T \boldsymbol{\beta}} e^{-e^{-\mathbf{x}^T \boldsymbol{\beta}} t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

2. (3 points) Consider a Weibull regression model with *exactly two* explanatory variables. If x_1 is increased by one unit, the hazard function at time t is multiplied by _____. That's the *hazard* function. Show your work, starting with a convenient expression on the formula sheet. You have more room than you need. **Circle your answer.**

$$\frac{1}{\sigma} \exp \left\{ -\frac{1}{\sigma} (\beta_0 + \beta_1 (x_1 + 1) + \beta_2 x_2) \right\} t^{\frac{1}{\sigma} - 1}$$

$$\frac{1}{\sigma} \exp \left\{ -\frac{1}{\sigma} (\beta_0 + \beta_1 x_1 + \beta_2 x_2) \right\} t^{\frac{1}{\sigma} - 1}$$

$$= e^{-\frac{1}{\sigma} (\beta_0 + \beta_1 x_1 + \beta_2 x_2)}$$

$$e^{-\frac{1}{\sigma} (\beta_0 + \beta_1 x_1 + \beta_2 x_2)}$$

$$= e^{-\beta_1/\sigma}$$

~~0 marks if wrong sign or missing~~
 -2 for wrong sign, -2 for missing σ , up to a maximum of -3

3. (3 points) In your analysis of the cancer data, you fit a Weibull regression model with just sex and physician's ECOG rating. You produced an estimate of median survival time for female patients with an ecog rating of 1, together with a 95% confidence interval. Write the following in the spaces provided.

(a) Estimated median survival time.

402.8872

(b) Lower 95% confidence limit.

323.2567

(c) Upper 95% confidence limit.

482.5177

On your printout, circle the three numbers and write "Question 3" beside them.

Please attach your printout. Make sure your name is on it.

R version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.79 (8198) x86_64-apple-darwin17.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

```
> # Assignment 8, Question 8
>
> rm(list=ls()); # options(scipen=999)
> # install.packages("survival",dependencies=TRUE) # Only need to do this once
> library(survival) # Do this every time
> # help(cancer)
> head(cancer)
  inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
1    3  306     2  74  1     1     90     100     1175     NA
2    3  455     2  68  1     0     90     90     1225     15
3    3 1010     1  56  1     0     90     90      NA     15
4    5  210     2  57  1     1     90     60     1150     11
5    1  883     2  60  1     0    100     90      NA      0
6   12 1022     1  74  1     1     50     80     513      0
> round( cor(cancer[,2:10],use='pairwise.complete.obs'), 3)
      time status  age  sex ph.ecog ph.karno pat.karno meal.cal wt.loss
time  1.000 -0.171 -0.078 0.130 -0.201  0.133  0.184  0.074  0.009
status -0.171 1.000  0.150 -0.243  0.233 -0.182 -0.172  0.024  0.035
age    -0.078 0.150 1.000 -0.122  0.193 -0.203 -0.126 -0.231  0.038
sex     0.130 -0.243 -0.122 1.000 -0.021  0.011  0.046 -0.168 -0.129
ph.ecog -0.201 0.233 0.193 -0.021 1.000 -0.807 -0.511 -0.099 0.188
ph.karno 0.133 -0.182 -0.203 0.011 -0.807 1.000  0.520  0.042 -0.175
pat.karno 0.184 -0.172 -0.126 0.046 -0.511 0.520 1.000  0.166 -0.172
meal.cal 0.074 0.024 -0.231 -0.168 -0.099 0.042 0.166 1.000 -0.103
wt.loss 0.009 0.035 0.038 -0.129 0.188 -0.175 -0.172 -0.103 1.000
>
> # Fix up variables
```

```

> Cancer = within(cancer,{
+ status=status-1 # Now status = delta
+ sex = sex-1      # So 0=M, 1=F
+ stime = Surv(time,status)
+ })
> summary(Cancer)

```

inst	time	status	age	sex
Min. : 1.00	Min. : 5.0	Min. : 0.0000	Min. : 39.00	Min. : 0.0000
1st Qu.: 3.00	1st Qu.: 166.8	1st Qu.: 0.0000	1st Qu.: 56.00	1st Qu.: 0.0000
Median : 11.00	Median : 255.5	Median : 1.0000	Median : 63.00	Median : 0.0000
Mean : 11.09	Mean : 305.2	Mean : 0.7237	Mean : 62.45	Mean : 0.3947
3rd Qu.: 16.00	3rd Qu.: 396.5	3rd Qu.: 1.0000	3rd Qu.: 69.00	3rd Qu.: 1.0000
Max. : 33.00	Max. : 1022.0	Max. : 1.0000	Max. : 82.00	Max. : 1.0000
NA's : 1				

ph.ecog	ph.karno	pat.karno	meal.cal
Min. : 0.0000	Min. : 50.00	Min. : 30.00	Min. : 96.0
1st Qu.: 0.0000	1st Qu.: 75.00	1st Qu.: 70.00	1st Qu.: 635.0
Median : 1.0000	Median : 80.00	Median : 80.00	Median : 975.0
Mean : 0.9515	Mean : 81.94	Mean : 79.96	Mean : 928.8
3rd Qu.: 1.0000	3rd Qu.: 90.00	3rd Qu.: 90.00	3rd Qu.: 1150.0
Max. : 3.0000	Max. : 100.00	Max. : 100.00	Max. : 2600.0
NA's : 1	NA's : 1	NA's : 3	NA's : 47

wt.loss	stime.time	stime.status
Min. : -24.000	Min. : 5.0000	Min. : 0.0000000
1st Qu.: 0.000	1st Qu.: 166.7500	1st Qu.: 0.0000000
Median : 7.000	Median : 255.5000	Median : 1.0000000
Mean : 9.832	Mean : 305.2325	Mean : 0.7236842
3rd Qu.: 15.750	3rd Qu.: 396.5000	3rd Qu.: 1.0000000
Max. : 68.000	Max. : 1022.0000	Max. : 1.0000000
NA's : 14		

```

>
> # (a)
>
> self = survreg(stime ~ pat.karno, dist='weibull', data=Cancer)
> summary(self)

```

Call:

```
survreg(formula = stime ~ pat.karno, data = Cancer, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	4.88007	0.31860	15.32	< 2e-16
pat.karno	0.01451	0.00399	3.63	0.00028
Log(scale)	-0.30890	0.06234	-4.95	7.2e-07

Scale= 0.734

Weibull distribution

Loglik(model)= -1127.9 Loglik(intercept only)= -1134.2

Chisq= 12.45 on 1 degrees of freedom, p= 0.00042

Number of Newton-Raphson Iterations: 5

n=225 (3 observations deleted due to missingness)

```

>
> # (b)
>
> full = survreg(stime ~ age + sex + ph.ecog + ph.karno + pat.karno + meal.cal +
wt.loss, dist='weibull', data=Cancer)
> summary(full)

```

Call:

```

survreg(formula = stime ~ age + sex + ph.ecog + ph.karno + pat.karno +
meal.cal + wt.loss, data = Cancer, dist = "weibull")

```

	Value	Std. Error	z	p
(Intercept)	7.38e+00	1.05e+00	7.03	2.1e-12
age	-6.42e-03	7.95e-03	-0.81	0.41983
sex	3.87e-01	1.41e-01	2.75	0.00593
ph.ecog	-5.21e-01	1.53e-01	-3.42	0.00063
ph.karno	-1.64e-02	7.54e-03	-2.17	0.02984
pat.karno	8.60e-03	5.53e-03	1.56	0.11979
meal.cal	-1.25e-05	1.78e-04	-0.07	0.94416
wt.loss	9.42e-03	5.32e-03	1.77	0.07631
Log(scale)	-3.63e-01	7.21e-02	-5.03	4.8e-07

Scale= 0.695

Weibull distribution

Loglik(model)= -833.4 Loglik(intercept only)= -847.6

Chisq= 28.48 on 7 degrees of freedom, p= 0.00018

Number of Newton-Raphson Iterations: 5

n=168 (60 observations deleted due to missingness)

```

>
> # (c)
>
> nodiet = survreg(stime ~ age + sex + ph.ecog + ph.karno + pat.karno,
dist='weibull', data=Cancer) # Dropping meal.cal and wt.loss
> anova(nodiet,full) # LR test

```

	Terms	Resid.	Df
1	age + sex + ph.ecog + ph.karno + pat.karno		216
2	age + sex + ph.ecog + ph.karno + pat.karno + meal.cal + wt.loss		159

	-2*LL	Test	Df	Deviance	Pr(>Chi)
1	2209.472	NA	NA	NA	NA
2	1666.763	=	57	542.709	2.301935e-80

```

> # Compare Wald test
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> # source("Wtest.txt") # Local copy
> Vnfull = vcov(full)
> Tnfull = full$coefficients; Tnfull = c(Tnfull,log(full$scale))
> L1 = rbind(c(0,0,0,0,0,0,1,0,0),
+           c(0,0,0,0,0,0,0,1,0) )
> Wtest(L=L1, Tn=Tnfull, Vn=Vnfull)
      W      df  p-value
3.1462651 2.0000000 0.2073945

```

```
>
> # (d)
>
> summary(nodiet)
```

Call:

```
survreg(formula = stime ~ age + sex + ph.ecog + ph.karno + pat.karno,
        data = Cancer, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	7.17935	0.88444	8.12	4.8e-16
age	-0.00764	0.00672	-1.14	0.2555
sex	0.39973	0.12155	3.29	0.0010
ph.ecog	-0.41366	0.13040	-3.17	0.0015
ph.karno	-0.01228	0.00684	-1.80	0.0725
pat.karno	0.00738	0.00484	1.52	0.1276
Log(scale)	-0.34464	0.06245	-5.52	3.4e-08

Scale= 0.708

Weibull distribution

Loglik(model)= -1104.7 Loglik(intercept only)= -1121.3

Chisq= 33.17 on 5 degrees of freedom, p= 3.5e-06

Number of Newton-Raphson Iterations: 5

n=223 (5 observations deleted due to missingness)

```
>
> # With missing values on the variables that are omitted from the restricted model,
statistics from the restricted model are based on a larger sample size. -2LL, badness
of fit, is a sum of negative terms. There are too many in the -2LL for the restricted
model, so the "badness" seems worse than it is.  $G^2 = -2LL$  for the restricted model
minus  $-2LL$  for the full model is too large, resulting in rejection of  $H_0$  too often.
This is a good example.
```

```
>
> # (e) Make a data frame with no missing values for the full model.
> # help(na.omit)
```

```
>
> dim(Cancer)
[1] 228 11
> Cancer2 = Cancer[,2:11]; dim(Cancer2) # Exclude column 1, institution
[1] 228 10
> Cancer2 = na.omit(Cancer2); dim(Cancer2)
[1] 168 10
```

```
>
>
> # (f)
>
> full2 = survreg(Surv(time,status) ~ age + sex + ph.ecog + ph.karno + pat.karno +
meal.cal + wt.loss, dist='weibull',data=Cancer2)
> nodiet2 = survreg(Surv(time,status) ~ age + sex + ph.ecog + ph.karno + pat.karno,
dist='weibull',data=Cancer2)
> anova(nodiet2,full2)
```



```

                                Terms Resid. Df
1                                age + sex + ph.ecog + ph.karno + pat.karno    161
2 age + sex + ph.ecog + ph.karno + pat.karno + meal.cal + wt.loss          159
-2*LL Test Df Deviance Pr(>Chi)
1 1670.043      NA      NA      NA
2 1666.763      = 2 3.279398 0.1940384
> # Note that update does not get the job done here. I might forget this point.
>
> # (g)
>
> model2 = survreg(stime ~ sex + ph.ecog + ph.karno, dist='weibull', data=Cancer);
summary(model2)

```

```

Call:
survreg(formula = stime ~ sex + ph.ecog + ph.karno, data = Cancer,
        dist = "weibull")

```

	Value	Std. Error	z	p
(Intercept)	7.07575	0.65093	10.87	< 2e-16
sex	0.41141	0.12287	3.35	0.00081
ph.ecog	-0.47729	0.12618	-3.78	0.00016
ph.karno	-0.00906	0.00675	-1.34	0.17983
Log(scale)	-0.32452	0.06207	-5.23	1.7e-07

Scale= 0.723

Weibull distribution

```

Loglik(model)= -1126.3  Loglik(intercept only)= -1141.1
Chisq= 29.49 on 3 degrees of freedom, p= 1.8e-06
Number of Newton-Raphson Iterations: 5
n=226 (2 observations deleted due to missingness)

```

```

> model3 = survreg(stime ~ sex + ph.ecog, dist='weibull', data=Cancer);
summary(model3)

```

```

Call:
survreg(formula = stime ~ sex + ph.ecog, data = Cancer, dist = "weibull")

```

	Value	Std. Error	z	p
(Intercept)	6.2210	0.1096	56.74	< 2e-16
sex	0.4014	0.1237	3.24	0.0012
ph.ecog	-0.3557	0.0826	-4.31	1.7e-05
Log(scale)	-0.3133	0.0613	-5.11	3.3e-07

Scale= 0.731

Weibull distribution

```

Loglik(model)= -1133.1  Loglik(intercept only)= -1147.4
Chisq= 28.73 on 2 degrees of freedom, p= 5.8e-07
Number of Newton-Raphson Iterations: 5
n=227 (1 observation deleted due to missingness)

```

```

> # model3.5 = survreg(stime ~ sex + ph.ecog + pat.karno, dist='weibull',

```

```

data=Cancer); summary(model3.5)
>
> # (g iv.)
> patient = data.frame(sex=1,ph.ecog=1)
> predlist = predict(model3,newdata=patient,type='quantile',p=0.5,se=TRUE)
> predlist
$fit
      1
402.8872
$se.fit
      1
40.62783

> mhat = predlist$fit; se = predlist$se
> L = mhat - 1.96*se; U = mhat+1.96*se
> c(L,U)
      1      1
323.2567 482.5177
>
> # Note ph.ecog=1 is the mode.
> table(Cancer$ph.ecog)

  0   1   2   3
63 113  50   1
>
>

```

Question 3