

Name Jenny

Student Number _____

STA 312 f2023 Quiz 10

1. (6 points) Prove $S(t) = e^{-\int_0^t h(y) dy}$. This is a fact from the formula sheet. You may use anything on the formula sheet *except* what you are proving.

$$\int_0^t h(y) dy = \int_0^t \frac{f(y)}{S(y)} dy = \int_0^t \frac{f(y)}{1-F(y)} dy$$

Let $u = 1 - F(y)$, $du = -f(y) dy$

y	$u = 1 - F(y)$
t	$1 - F(t) = S(t)$

$$= - \int_{S(t)}^1 \frac{1}{u} du = \int_{S(t)}^1 \frac{1}{u} du$$

$$= \log u \Big|_{S(t)}^1 = \log(1) - \log(S(t)) = -\log S(t)$$

So,

$$-\log S(t) = \int_0^t h(y) dy \Rightarrow \log S(t) = - \int_0^t h(y) dy$$

$$\Rightarrow S(t) = e^{-\int_0^t h(y) dy}$$

2. In your analysis of the veteran data, you fit a model with just experimental treatment, cell type (type of cancer) and Karnofsky score. You found that controlling for experimental treatment and Karnofsky score, the risk of death did depend on cell type. Following that finding up with pairwise comparisons, you tested for a difference between adenocarcinoma and large cell cancer.

(a) (1 point) Fill in the table below.

Z or Chi-squared Statistic (a number)	p-value (a number)	Reject H_0 ? (Yes or No)	Statistically Significant? (Yes or No)
$W = 6.46$	$p = 0.011$	Yes	Yes

On your printout, circle the test statistic value and write "Question 2" beside it.

(b) (2 points) State the conclusion in plain, non-statistical language. Use the word "adjusting."

Adjusting for experimental treatment and Karnofsky score, the risk of death from adeno is greater than the risk of death from large cell cancer.

No marks unless the test statistic and rejection of H_0 is correct in (a).

One mark off for a non-directional conclusion.

Zero if "p-value", "null hypothesis" etc.

(c) (1 point) Allowing for Karnofsky score and treatment, the risk of death for a patient with adenocarcinoma is estimated to be _____ times as great as the risk for a patient with large cell cancer.

$$e^{1.153994 - 0.394625} = e^{0.759} = 2.14$$

Please attach your printout for the veteran data. 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]

```
> # ----- veteran ----- #
```

```
>  
> rm(list=ls()); options(scipen=999)  
> library(survival)  
> help(veteran)  
starting httpd help server ... done  
> summary(veteran)
```

trt	celltype	time	status	karno
Min. :1.000	squamous :35	Min. : 1.0	Min. :0.0000	Min. :10.00
1st Qu.:1.000	smallcell:48	1st Qu.: 25.0	1st Qu.:1.0000	1st Qu.:40.00
Median :1.000	adeno :27	Median : 80.0	Median :1.0000	Median :60.00
Mean :1.496	large :27	Mean :121.6	Mean :0.9343	Mean :58.57
3rd Qu.:2.000		3rd Qu.:144.0	3rd Qu.:1.0000	3rd Qu.:75.00
Max. :2.000		Max. :999.0	Max. :1.0000	Max. :99.00

diagtime	age	prior
Min. : 1.000	Min. :34.00	Min. : 0.00
1st Qu.: 3.000	1st Qu.:51.00	1st Qu.: 0.00
Median : 5.000	Median :62.00	Median : 0.00
Mean : 8.774	Mean :58.31	Mean : 2.92
3rd Qu.:11.000	3rd Qu.:66.00	3rd Qu.:10.00
Max. :87.000	Max. :81.00	Max. :10.00

```
> head(veteran)
```

	trt	celltype	time	status	karno	diagtime	age	prior
1	1	squamous	72	1	60	7	69	0
2	1	squamous	411	1	70	5	64	10
3	1	squamous	228	1	60	3	38	0
4	1	squamous	126	1	60	9	63	10
5	1	squamous	118	1	70	11	65	10
6	1	squamous	10	1	20	5	49	0

```

>
>
> contrasts(veteran$celltype)
      smallcell adeno large
squamous      0      0      0
smallcell      1      0      0
adeno          0      1      0
large          0      0      1
> # Edit data frame
> vet = within(veteran,
+ {
+ trt = trt - 1 # Makes it indicator for test treatment; standard is reference.
+ prior = prior/10 # Now 1=yes.
+ })
>
> # Experimental treatment, cell type and Karnofsky score
> model1 = coxph(Surv(time,status) ~ trt + celltype + karno, data=vet); summary(model1)
Call:
coxph(formula = Surv(time, status) ~ trt + celltype + karno,
      data = vet)

      n= 137, number of events= 128

              coef exp(coef) se(coef)      z      Pr(>|z|)
trt          0.261744  1.299194  0.200923  1.303      0.19267
celltypesmallcell 0.824980  2.281836  0.268911  3.068      0.00216 **
celltypeadeno    1.153994  3.170833  0.295038  3.911 0.00009178351 ***
celltypelarge    0.394625  1.483828  0.282243  1.398      0.16206
karno           -0.031271  0.969213  0.005165 -6.054 0.00000000141 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
trt              1.2992    0.7697    0.8763    1.9262
celltypesmallcell 2.2818    0.4382    1.3471    3.8653
celltypeadeno    3.1708    0.3154    1.7784    5.6534
celltypelarge    1.4838    0.6739    0.8534    2.5801
karno             0.9692    1.0318    0.9595    0.9791

Concordance= 0.737 (se = 0.022 )
Likelihood ratio test= 61.07 on 5 df,  p=0.000000000007
Wald test              = 63.41 on 5 df,  p=0.000000000002
Score (logrank) test = 66.55 on 5 df,  p=0.000000000005

>
> # If treatment is left alone,
> summary(coxph(Surv(time,status) ~ trt + celltype + karno, data=veteran))
Call:
coxph(formula = Surv(time, status) ~ trt + celltype + karno,
      data = veteran)

      n= 137, number of events= 128

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
trt	0.261744	1.299194	0.200923	1.303	0.19267
celltypesmallcell	0.824980	2.281836	0.268911	3.068	0.00216 **
celltypeadeno	1.153994	3.170833	0.295038	3.911	0.00009178351 ***
celltypelarge	0.394625	1.483828	0.282243	1.398	0.16206
karno	-0.031271	0.969213	0.005165	-6.054	0.00000000141 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
trt	1.2992	0.7697	0.8763	1.9262
celltypesmallcell	2.2818	0.4382	1.3471	3.8653
celltypeadeno	3.1708	0.3154	1.7784	5.6534
celltypelarge	1.4838	0.6739	0.8534	2.5801
karno	0.9692	1.0318	0.9595	0.9791

Concordance= 0.737 (se = 0.022)
Likelihood ratio test= 61.07 on 5 df, p=0.000000000007
Wald test = 63.41 on 5 df, p=0.000000000002
Score (logrank) test = 66.55 on 5 df, p=0.000000000005

```

>
> nocelltype = update(model1, . ~ .-celltype)
> anova(nocelltype,model1)
Analysis of Deviance Table
Cox model: response is Surv(time, status)
Model 1: ~ trt + karno
Model 2: ~ trt + celltype + karno
  loglik  Chisq Df Pr(>|Chi|)
1 -483.97
2 -474.91 18.102 3 0.000419 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> # Pairwise comparisons (use Wald tests)
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> Vn1 = vcov(model1); b1 = coefficients(model1)
> smallVSadeno = cbind(0, 1,-1, 0, 0)
> smallVSlarge = cbind(0, 1, 0,-1, 0)
> adenoVSlarge = cbind(0, 0, 1,-1, 0)
> Wtest(L=smallVSadeno, Tn=b1, Vn=Vn1)
      W      df  p-value
1.4953326 1.0000000 0.2213909
> Wtest(L=smallVSlarge, Tn=b1, Vn=Vn1)
      W      df  p-value
2.6803131 1.0000000 0.1015957
> Wtest(L=adenoVSlarge, Tn=b1, Vn=Vn1) # Quiz question on this one. Interpret.
      W      df  p-value
6.46489623 1.0000000 0.01100261
>
>
>
>

```

Question 2

```

>
> # ----- cancer again ----- #
>
> 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)
> # summary(cancer)
>
> attach(cancer)
> status=status-1; sex = sex-1 # So 0=M, 1=F
> stime = Surv(time,status)
> table(ph.ecog)
ph.ecog
  0  1  2  3
63 113 50  1
> summary(ph.ecog)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
0.0000 0.0000  1.0000  0.9515  1.0000  3.0000    1
> # Eliminate the ph.ecog=3
> ecog = ph.ecog
> ecog[ecog==3] = NA
> table(ecog) # 0, 1, 2
ecog
  0  1  2
63 113 50
> ecog = factor(ecog) # This makes zero (good) the ref category -- okay.
>
> model1 = coxph(stime ~ sex + ecog); summary(model1)
Call:
coxph(formula = stime ~ sex + ecog)

n= 226, number of events= 163
(2 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
sex    -0.5447    0.5800   0.1681 -3.240   0.0012 **
ecog1   0.4180    1.5189   0.1995  2.096   0.0361 *
ecog2   0.9464    2.5765   0.2248  4.211 0.0000254 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
sex            0.580    1.7242    0.4172    0.8064
ecog1          1.519    0.6584    1.0274    2.2454
ecog2          2.577    0.3881    1.6585    4.0027

Concordance= 0.64 (se = 0.025 )
Likelihood ratio test= 27.58 on 3 df,  p=0.000004
Wald test              = 27.85 on 3 df,  p=0.000004
Score (logrank) test = 28.82 on 3 df,  p=0.000002

> # Test whether hazard ratio of 1 to 0 = hr of 2 to 1

```

```
> # H0: beta2 = beta3-beta2 <=> 2 beta2 = beta3
> LL = cbind(0,2,-1)
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> betahat = coefficients(model1); kov = vcov(model1)
> Wtest(L=LL, Tn=betahat, Vn=kov)
      W      df  p-value
0.1229529 1.0000000 0.7258537
>
> # ----- #
>
```