Name Jerry

Student Number

STA 312 s2019 Quiz 9

1. (4 points) Show that if $\widehat{\boldsymbol{\theta}}_n \dot{\sim} N_k(\boldsymbol{\theta}, \mathbf{V}_n)$ and $\mathbf{a}$ is a non-zero $k \times 1$ vector of constants, then $W = \mathbf{a}^\top \widehat{\boldsymbol{\theta}}_n \dot{\sim} N\left(\mathbf{a}^\top \boldsymbol{\theta}, \mathbf{a}^\top \mathbf{V}_n \mathbf{a}\right)$

$g(\theta) = a^T\theta = a_1\theta_1 + a_2\theta_2 + \cdots + a_k\theta_k$

$\dot{g}(\theta) = \left(\frac{\partial g}{\partial \theta_1}, \frac{\partial g}{\partial \theta_2}, \cdots, \frac{\partial g}{\partial \theta_k}\right)$

$= (a_1, a_2, \cdots a_k) = a^T$

So by the delta method on the formula sheet

$a^T\hat{\theta}_n = g(\hat{\theta}_n) \dot{\sim} N(g(\theta), \dot{g}(\theta) V_n \dot{g}(\theta)^T)$

$= N(a^T\theta, a^T V_n a)$

2. In your analysis of the **cancer** data, you fit a log-normal regression model with just sex and physician's ECOG rating.

   (a) (2 points) Look at the test of physician's ECOG rating controlling for sex. What is it telling you? Do patients with a higher ECOG rating tend to survive longer, or do they tend to live a shorter length of time?

Patients with a higher ECOG rating tend to live a shorter time.

   (b) (2 points) Allowing for gender, the median survival time for a patient with an ECOG rating of zero is estimated to be _____ times as long as the median survival time for a patient with an ECOG rating of one. Write the answer in the space below. Use a calculator if you have one. *Your answer must be consistent with your answer to the last question.*

$1/e^{-0.3936} = e^{0.3936} = 1.48$

↑ This is okay if no calculator

IF (a) & (b) ARE INCONSISTENT, THEY ARE BOTH WRONG

   (c) (2 points) Give a 95% prediction interval for the survival time (in days) for a new female patient with an ECOG rating of one. Your answer is a pair of numbers. Write the numbers in the space below. On your printout, circle the two numbers and write "Question 2c" beside them.

(52, 3175)

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

```
> # A9: Log-Normal
> 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)
> full = survreg(stime ~ age + sex + ph.ecog + ph.karno + pat.karno + meal.cal + wt.loss,
dist='lognormal')
> summary(full)

Call:
survreg(formula = stime ~ age + sex + ph.ecog + ph.karno + pat.karno +
    meal.cal + wt.loss, dist = "lognormal")
               Value Std. Error       z          p
(Intercept)  6.750192   1.463466  4.6125 0.00000398
age         -0.013269   0.010130 -1.3099 0.19022275
sex          0.478505   0.181937  2.6301 0.00853681
ph.ecog     -0.465406   0.205081 -2.2694 0.02324544
ph.karno    -0.011976   0.011706 -1.0231 0.30627702
pat.karno    0.008712   0.006702  1.2999 0.19362309
meal.cal     0.000279   0.000218  1.2812 0.20012270
wt.loss      0.008169   0.006302  1.2961 0.19493560
Log(scale)   0.002878   0.065132  0.0442 0.96475331

Scale= 1

Log Normal distribution
Loglik(model)= -846.1   Loglik(intercept only)= -860.2
    Chisq= 28.22 on 7 degrees of freedom, p= 0.0002
Number of Newton-Raphson Iterations: 4
n=168 (60 observations deleted due to missingness)

>
> # Test other variables controlling for sex and ph.ecog.
>
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> Vnfull = vcov(full)
> Tnfull = full$coefficients; Tnfull = c(Tnfull,log(full$scale))
> LL = rbind( c(0,1,0,0,0,0,0,0,0),
+             c(0,0,0,0,1,0,0,0,0),
+             c(0,0,0,0,0,1,0,0,0),
+             c(0,0,0,0,0,0,1,0,0),
+             c(0,0,0,0,0,0,0,1,0)   )
> Wtest(L=LL, Tn=Tnfull, Vn=Vnfull)
        W        df   p-value
8.6372440 5.0000000 0.1244379
>
> model2 = survreg(stime ~ sex + ph.ecog, dist='lognormal'); summary(model2)

Call:
survreg(formula = stime ~ sex + ph.ecog, dist = "lognormal")
              Value Std. Error      z        p
(Intercept)  5.8366      0.138 42.354 0.000000
sex          0.5641      0.154  3.673 0.000239
ph.ecog     -0.3936      0.103 -3.808 0.000140
Log(scale)   0.0407      0.056  0.727 0.467267

Scale= 1.04

Log Normal distribution
Loglik(model)= -1149.5   Loglik(intercept only)= -1163.2
    Chisq= 27.3 on 2 degrees of freedom, p= 0.0000012
Number of Newton-Raphson Iterations: 3
n=227 (1 observation deleted due to missingness)
```

```
> exp(0.5641)
[1] 1.757865
>
> # Prediction and CI
> fem1 = data.frame(sex=1,ph.ecog=1)
> pred = predict(model2,newdata=fem1,type='linear',se=TRUE) ; pred
$fit
       1
6.007115

$se.fit
        1
0.1242394

> yhat = pred$fit
> that= exp(yhat)
> that # Prediction = estimated median
       1
406.3095
>
> # Prediction interval
> sigmasqhat = model2$scale^2
> se = sqrt(sigmasqhat+pred$se^2)
> L = yhat - 1.96*se; U = yhat + 1.96*se
> lower95 = exp(L); upper95 = exp(U)
> predint = c(that,lower95,upper95)
> names(predint) = c('t-hat','lower95','upper95')
> predint
     t-hat    lower95    upper95
 406.30954  51.99567 3175.02309
>
> # Confidence interval for median
> Lci = yhat - 1.96*pred$se; Uci = yhat + 1.96*pred$se
> lower95ci = exp(Lci); upper95ci = exp(Uci)
> ci = c(that,lower95ci,upper95ci)
> names(ci) = c('t-hat','lower95ci','upper95ci')
> ci
    t-hat lower95ci upper95ci
 406.3095  318.4948  518.3364
>
```

Question 2c