ADVANCED STATISTICAL METHODOLOGY (STAT 526)
FINAL EXAM (SC 239)
3:30pm-5:30PM, Wednesday, December 12, 2018

There are totally 38 points in the exam. The students with score higher than or equal to 35 points will receive 35 points. Please write down your name and student ID number below.

NAME: ______________________________

ID: ______________________________

1. (10 points). The data set reports the result of an experiment on radiation of rats. Eight radiation amount values (i.e., $0, 1, \cdots, 7$) were selected with a number of rats at each value. The statuses of the results were given by normal (status = 1), sick (status = 2), and death (status = 3). The counts of rats at each radiation amount value with respect to the three statuses were collected. The R output is given below.

```
> gg0 <- multinom(status~factor(radiation),weights=count)
> gg0$deviance
[1] 721.2502
> gg1 <- multinom(status~radiation,weights=count)
> gg1$deviance
[1] 733.4818
> summary(gg1)
Call:
multinom(formula = status ~ radiation, weights = count)
Coefficients:
  (Intercept) radiation
2  -0.03740    0.4594
3  -0.34579    0.4796
> gg2 <- polr(status~radiation,weights=count)
> summary(gg2)
Call:
polr(formula = status ~ radiation, weights = count)
Coefficients:
           Value Std. Error t value
radiation 0.1957      0.043   4.553
Intercepts:
    Value   Std. Error t value
1|2 -1.0933  0.1914    -5.7124
2|3  1.2285  0.1920     6.4002
Residual Deviance: 755.9612
```

(a) (2 points). State model assumptions of the multinomial and proportional odds models when radiation amount is treated as a continuous variable. Provide the estimates of model parameters.

(b) (2 points). Provide residual deviance goodness-of-fit statistics to test whether the two models in the previous part fit the data, respectively. You need to state the values of the test statistics, their degrees of freedom, and the conclusion.

(c) (2 points). Compute the fitted probabilities under the multinomial and proportional odds models of the three statuses when radiation equals 6, respectively.

(d) (2 points). Compute the odds ratio between normal and death under the multinomial model when radiation changes from 5 to 6.

(e) (2 points). Compute the odds ratio between normal and death under the proportional odds model when radiation changes from 5 to 6.

2. (8 points). The data set reports the strength of a kind of metal made by four different methods. Each method made 5 samples. The sample averages of the data are $\bar{y}_{1\cdot} = 4.3473$, $\bar{y}_{2\cdot} = 3.2154$, $\bar{y}_{3\cdot} = 1.8663$, and $\bar{y}_{4\cdot} = 3.1610$. The MSE of the data is $\hat{\sigma}^2 = 1.7362$.

   (a) (2 points). Complete the following ANOVA table.

Source	DF	SS	MS	F-Value
Method				
Error				
Total				

   (b) (2 points). State the assmptions of the one-way random effects model and provide the estimate values of the model parameters.

   (c) (2 points). State the model and provide estimates of parameters in the model when the random effects are not included.

   (d) (2 points). Describe a bootstrap test to assess whether the random effects can be ignored.

3. (10 points). The dataset reports the arm strength measured on 8 patients at 3 different times and where patients were randomized to one of 2 treatment groups. The R output is given.

```
Linear mixed-effects model fit by REML
 Data: strength
Random effects:
 Formula: ~time | factor(Subject)
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
(Intercept) 1.441144 (Intr)
time        1.271301 -0.567
Residual    0.790521
Fixed effects: Strength ~ factor(treatment) + time
                      Value Std.Error DF   t-value p-value
(Intercept)         7.785315 0.7080676 15 10.995158  <.0001
factor(treatment)2 -4.070631 0.9288906  6 -4.382250  0.0047
time                2.000000 0.4910027 15  4.073297  0.0010
 Correlation:
                   (Intr) fct()2
factor(treatment)2 -0.656
time               -0.374  0.000
```

(a) (2 points). Write down the model assumptions of the output.

(b) (2 points). Write down the fitted model with respect to the two different treatment groups.

(c) (2 points). Explain why the random effect term must be added in the model.

(d) (2 points). Compute the 95% confidence interval for the mean response when time equal to $t = 2.5$ in the first treatment group.

(e) (2 points). Compute the 95% confidence interval for the observed value of the response when time equal to $t = 2.5$ in the first treatment group.

4. (10 points). The data reports the survival time in month for 42 patients with leukemia with respect to 3 different treatment methods with death indicator (dropoff(0) and death(1)). Let **trt** be 1 for placebo, 2 for new medicine, and 3 for old medicine, respectively.

Treatment	Suvival Time
Placebo	1 1+ 2 4 5+ 7 10 13 13+ 17 20 25 27+ 29
New Medicine	7 8+ 12 18+ 22 20 21 27+ 29 30+ 39 42 45+ 49
Old Medicine	3 3 4 7+ 11 17+ 18 22+ 29 30+ 32 34+ 38 40

```
Call: survdiff(formula = Surv(month, censor) ~ factor(trt))
              N Observed Expected (O-E)^2/E (O-E)^2/V
factor(trt)=1 14       10     4.84   5.51279    7.4751
factor(trt)=2 14        9    13.96   1.76165    4.1350
factor(trt)=3 14        9     9.20   0.00455    0.0074
 Chisq= 8.4  on 2 degrees of freedom, p= 0.0147
Call: survreg(formula = Surv(month, censor) ~ factor(trt), dist = "weibull")
              Value Std. Error     z        p
(Intercept)   2.876      0.214 13.41 5.10e-41
factor(trt)2  0.751      0.313  2.40 1.64e-02
factor(trt)3  0.540      0.312  1.73 8.31e-02
Log(scale)   -0.392      0.159 -2.47 1.36e-02
Call: coxph(formula = Surv(month, censor) ~ factor(trt))
                coef exp(coef) se(coef)      z Pr(>|z|)
factor(trt)2 -1.4730    0.2292   0.5410 -2.722  0.00648 **
factor(trt)3 -0.9093    0.4028   0.4913 -1.851  0.06421 .
```

(a) (2 points). Compute the Kaplan-Meier estimator of the survival function for the placebo group for $t \leq 10$. Test whether the survival functions are all equal.

(b) (2 points). Estimate the hazard functions for all the treatment groups if the survival functions are assumed exponentially distributed.

(c) (2 points). Derive the estimates of the hazard functions for all the groups under Weibull distribution. Can they be reduced to an exponential distribution.

(d) (2 points). Write down the assumption of the Cox proportional hazard model and derive the estimate of parameters.

(e) (2 points). Do you think the survival functions are equal under the model assumption of part (d). Explain.