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.

Solution: Let x be the radiation amount, $\pi_1(x)$, $\pi_2(x)$, $\pi_3(x)$ be the probabilities of normal, sick, and death, respectively. The baseline multinomial model is

$$\log \frac{\pi_2(x)}{\pi_1(x)} = \beta_{01} + \beta_{11}x$$

and

$$\log \frac{\pi_3(x)}{\pi_1(x)} = \beta_{02} + \beta_{12}x.$$

The estimates of these parameters are $\hat{\beta}_{01} = -0.03470$, $\hat{\beta}_{11} = 0.4594$, $\hat{\beta}_{02} = -0.34579$, and $\hat{\beta}_{12} = 0.4796$. The proportional odds model is

$$\log \frac{\pi_1(x)}{\pi_2(x) + \pi_3(x)} = \eta_1 - \beta x$$

and

$$\log \frac{\pi_1(x) + \pi_2(x)}{\pi_3(x)} = \eta_2 - \beta x.$$

The estimates of these parameters are $\hat{\eta}_1 = -1.0933$, $\hat{\eta}_2 = 1.2285$, and $\hat{\beta} = 0.1957$.

(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.

Solution: The residual deviance G^2 of the multinomial model is $733.4818 - 721.2502 = 12.23$ with $16 - 4 = 12$ degrees of freedom. Since it is less than $\chi^2_{0.05,12} = 21.03$, we conclude that the model fits the data. The residual deviance G^2 of the proportional odds model is $755.9612 - 721.2502 = 34.721$ with $16 - 3 = 13$ degrees of freedom. SInce it is greater than $\chi^2_{0.05,13} = 22.36$, we conclude that the proportional odds model does not fit the data.

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

Solution: By the output of the multinomial model, we have $\pi_2(6)/\pi_1(6) = \exp(-0.03740 + 0.4594(6)) = 15.16$, $\pi_3(6)/\pi_1(6) = \exp(-0.34579 + 0.4796(6)) - 12.57$. Together with $\pi_1(6) + \pi_2(6) + \pi_3(6) = 1$, we have $\pi_1(6) = 0.0348$, $\pi_2(6) = 0.5276$, and $\pi_3(x) = 0.4376$. By the output of the proportional odds model, we have $\pi_1(6) = \exp(-1.0933 - 0.1957(6))/[1 + \exp(-1.0933 - 0.1957(6))] = 0.0939$ and $\pi_3(6) = 1/[1 + \exp(1.2285 - 0.1957(6))] = 0.4864$, leading to $\pi_2(6) = 1 - 0.0393 - 0.4864 = 0.4743$.

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

Solution: Using the same method in the previous part when radiation equals 5, we have $\pi_1(5) = 0.0544$, $\pi_2(5) = 0.5216$, and $\pi_2(5) = 0.4239$. Thus, the odds ratio is $\hat{\theta} = \pi_1(5)\pi_3(6)/[\pi_1(6)\pi_3(5)] = 0.0544 \times 0.4376/(0.0348 \times 0.4239) = 1.6137$.

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

Solution: Using the same method to the proportional odds model, we have $\pi_1(5) = 0.1118$, $\pi_2(5) = 0.4503$, and $\pi_3(5) = 0.4378$. We have $\hat{\theta} = 0.1118 \times 0.4864/(0.0939 \times 0.4378) = 1.3228$.

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				

Solution: We need to compute SSA. By $\bar{y}_{..} = (4.3473 + 3.2154 + 1.8663 + 3.1610)/4 = 3.1475$. We have

$$\begin{aligned} SSA =& 5[(4.3473 - 3.1475)^2 + (3.2154 - 3.1475)^2 \\ & + (1.8663 - 3.1475)^2 + (3.1610 - 3.1475)^2] = 15.4289. \end{aligned}$$

Then, we have the following table.

Source	DF	SS	MS	F-Value
Method	3	15.4289	5.1430	2.9623
Error	16	27.7792	1.7362	
Total	19	43.2081		

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

Solution: Let y_{ij} be the jth value at the ith method for $i \in \{1, 2, 3, 4\}$ and $j \in \{1, 2, 3, 4, 5\}$. The one-way random effects model can be expressed as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where $\alpha_i \sim^{iid} N(0, \sigma_\alpha^2)$ and $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$ independently. By the table in the previous part, we have $\hat{\sigma}^2 = 1.7362$, and $\hat{\sigma}_\alpha^2 = (5.1430 - 1.7362)/5 = 0.6814$.

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

Solution: If random effects are not included, then the model becomes

$$y_{ij} = \mu + \epsilon_{ij},$$

where $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$. The estimates are $\hat{\mu} = 3.1475$ and $\hat{\sigma}^2 = (15.4289+27.7792)/19 = 2.2741$.

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

Solution: We test the hypotheses

$$H_0 : \sigma_\alpha^2 = 0 \leftrightarrow H_1 : \sigma_\alpha^2 > 0.$$

At the beginning, we calculate the likelihood ratio statistic between models in (b) and (c) using the observed data. After that, we generate the response y_{ij} independently from $N(3.1475, 2.22711)$. We use the simulated data to compute the likelihood ratio statistic. We repeat the simulation 1000 times. We use those to derive the bootstrap distribution of the test statistic. The p-value of the observed likelihood ratio statistic based on the bootstrap distribution. We reject H_0 if the observed likeihood ratio statistic is significantly large.

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.

Solution: The model assumption is

$$y = \beta_0 + \tau + time\beta_1 + \gamma_0 + time\gamma_1 + \epsilon,$$

where β_0, β_1, τ are fixed effect, $\tau = 1$ if treatment is 2 and $\tau = 0$ if treatment is 0; γ_0 and γ_1 are random effects and ϵ is an error term. We assume $\gamma_0 \sim N(0, \sigma_0^2)$, $\gamma_1 \sim N(0, \sigma_1^2)$, $Cov(\gamma_0, \gamma_1) = \sigma_{01}$, and $\epsilon \sim N(0, \sigma^2)$; the random effect and the error term are independent but the random effect may not.

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

Solution: The fitted model for treatment 1 is

$$y = b_0 + b_1 time$$

where $b_0 \sim N(7.7853, 1.44^2)$ and $b_1 \sim N(2, 1.27^2)$. The fitted model for treatment 2 is

$$y = b_0 + b_1 time$$

where $b_0 \sim N(3.7147, 1.44^2)$ and $b_2 \sim N(2, 1.27^2)$.

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

Solution: Since the recovery of the arm strength depends on patients themselves, the random effect describes the difference and must be included 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.

Solution: The estimated value is

$$7.7853 + 2.5 \times 2 = 12.7853.$$

Its variance is

$$\begin{pmatrix} 1 & 2.5 \end{pmatrix} \begin{pmatrix} 0.7081^2 & -0.374 \times 0.7081 \times 0.4910 \\ * & 0.4910^2 \end{pmatrix} \begin{pmatrix} 1 \\ 2.5 \end{pmatrix} = 1.3580.$$

Therefore, the 95% confidence interval is

$$12.7853 \pm 1.96\sqrt{1.3580} = [10.5013, 15.0694].$$

(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.

Solution: The variance of the random effect term is

$$\begin{pmatrix} 1 & 2.5 \end{pmatrix} \begin{pmatrix} 1.4411^2 & -0.567 \times 1.4411 \times 1.2713 \\ * & 1.2713^2 \end{pmatrix} \begin{pmatrix} 1 \\ 2.5 \end{pmatrix} = 6.9841.$$

Therefore, the 95% confidence interval is

$$12.7853 \pm 1.96\sqrt{1.3580 + 6.9841 + 0.7905^2} = [6.9161, 18.6545].$$

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.

Solution: The estimate of the survival function is $\hat{S}(1) = (1 - 1/14) = 0.9285$, $\hat{S}(2) = 0.9286(1 - 1/12) = 0.8512$, $\hat{S}(4) = 0.882(1 - 1/11) = 0.7738$, $\hat{S}(7) = 0.7738(1 - 1/9) = 0.6878$, $\hat{S}(10) = 0.6878(1 - 1/8) = 0.6019$, $\hat{S}(13) = 0.6018(1 - 1/7) = 0.5159$, $\hat{S}(20) = 0.5159(1 - 1/5) = 0.4127$, $\hat{S}(25) = 0.4127(1 - 1/3) = 0.2751$, $\hat{S}(29) = 0$.

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

Solution: Using $\hat{\lambda} = \sum_{i=1}^{n} \delta_i / \sum_{i=1}^{n} t_i$, we obtain $\hat{\lambda}_1 = 0.05747$, $\hat{\lambda}_2 = 0.02439$, and $\hat{\lambda}_3 = 0.03125$.

(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.

Solution: Using the formula $h(t) = \alpha t^{\alpha-1} e^{-\beta x}$, we obtain $h_1(t) = 0.03808 t^{-0.3243}$, $h_2(t) = 0.01797 t^{-0.3243}$, and $h_3(t) = 0.02219 t^{-0.3243}$.

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

Solution: The model assumption is $h_j(t) = h_1(t) e^{\beta_j}$ for $j = 2, 3$. Their estimates are $\hat{\beta}_2 = -1.4730$ and $\hat{\beta}_3 = -0.9093$.

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

Solution: The survival functions are not all equal because the p-value of β_2 is significant.