STAT 410 SECOND MIDTERM EXAMINATION SOLUTION

Problem 1. Dermatologists in a hospital study patients with acute psoriasis, a skin disease. They would like to know whether medication A is more effective in relieving the symptoms of psoriasis than medication B. The data are retrospectively collected on 70 patients. The variables are gender (M/F), age (in years), medication(A/B), and response (1=relief, 0=no relief). Use the code and output to answer the questions below.

(a) What models were fitted to these data? What model has the best fit? Explain by giving a quantitative answer.

Criterion	Binary Logistic	Binary Probit	Binary cloglog
AIC	43.5408	44.7893	54.6158
AICC	45.1408	46.3893	56.2158
BIC	49.1456	50.3941	60.2206

The smallest values in all three criteria are for the binary logistic model, so it has the best fit.

(b) Write down the best-fitted model. Estimate all parameters.

The fitted logistic model can be written as

$$\frac{\hat{\pi}}{1-\hat{\pi}} = \exp(-3.2544 + 0.9640 \cdot male + 0.0780 \cdot age + 1.5574 \cdot medA).$$

(c) Interpret only the estimated regression coefficients that are significant at the 5% level.

The significant at the 5% level are gender and age. For males, the estimated odds in favor of relief from psoriasis are $\exp(0.9640) \cdot 100\% = 262.22\%$ of those for females. As age increases by one year, the estimated odds of psoriasis relief increase by $(\exp(0.0780) - 1) \cdot 100\% = 8.11\%$.

(d) Use the best-fitted model for prediction. Find the predicted probability of relief from psoriasis for a 32-year-old male taking medication B. Compute the value by hand and compare it to the one outputted by SAS.

$$\pi^0 = \frac{\exp(-3.2544 + 0.9640 \cdot 1 + 0.0780 \cdot 32)}{1 + \exp(-3.2544 + 0.9640 \cdot 1 + 0.0780 \cdot 32)} = 0.55122,$$

in SAS $\pi^0 = 0.55138$.

Problem 2. A satellite television provider is focusing on improving customer service. The company surveys subscribers who contact the call center and records how long the callers have been subscribed to the company (in months), whether they receive their monthly programming magazine (yes/no), whether the issue they called about was resolved (yes/no), and the overall satisfaction with the customer service, measured on a 5-point Likert scale (1=very dissatisfied, 2=dissatisfied, 3=neutral, 4=satisfied, 5=very satisfied). Use the code and output to answer the following questions:

(a) Which model fits the data the best? Give a quantitative answer.

Criterion	Cumulative logit	Cumulative probit	Cumulative clog-log
AIC	116.83	117.58	115.94
AICC	120.8304	121.5807	119.9387
BIC	127.9151	128.6653	127.0234

The smallest values in all three criteria are for the cumulative complementary log-log model, so it has the best fit.

(b) For the model chosen in part (a), write down the fitted model. Estimate all parameters.

The fitted model is

- $\hat{P}(satisf = 1) = \hat{P}(very\ dissatisfied) = 1 - \exp(-\exp(-3.4666 - 0.013007 \cdot \#of\ months\ subscribed + 1.451553\ \cdot no\ magazine + 0.85081 \cdot issue\ not\ resolved))$, $\hat{P}(very\ dissatisfied\ or\ dissatisfied) = 1 - \exp(-\exp(-1.9716 - 0.013007 \cdot \#of\ months\ subscribed + 1.451553 \cdot no\ magazine + 0.85081 \cdot issue\ not\ resolved))$,
- $\hat{P}(very\ dissatisfied, dissatisfied, or\ neutral) = 1 - \exp(-\exp(-0.5814 - 0.013007 \cdot \#of\ months\ subscribed + 1.451553 \cdot no\ magazine + 0.85081 \cdot issue\ not\ resolved))$,

  and
- $\hat{P}(very\ dissatisfied, dissatistied, neutral, or\ satisfied) = 1 - \exp(-\exp(0.8883 - 0.013007 \cdot \#of\ months\ subscribed + 1.451553 \cdot no\ magazine + 0.85081 \cdot issue\ not\ resolved))$.

(c) What predictors are significant at the 5% level? Give an interpretation of the estimated significant regression coefficients.

Significant predictor is a subscription to the magazine. The estimated complementary probabilities for customers who don't receive the magazine are those for customers who receive the magazine raised to the power $\exp(1.451553) = 4.26974$. It means that customers who don't receive the magazine are much less satisfied than those who receive the magazine (the probability of more satisfaction is much smaller).

(d) Predict probabilities of each of the 5 levels of the satisfaction score for a caller who has been subscribed for 8 months, doesn't receive the magazine, and whose issue was resolved over the phone. Compare your results with the R output.

The cumulative predicted probabilities are: $P^0(very\ dissatisfied) = 1 - \exp(-\exp(-3.4666 - 0.013007 \cdot 8 + 1.451553)) = 0.113203$,

$P^0(very\ dissatisfied\ or\ dissatisfied) = 1 - \exp(-\exp(-1.9716 - 0.013007 \cdot 8 + 1.451553)) = 0.414765$,

$P^0(very\ dissatisfied, dissatisfied, or\ neutral) = 1 - \exp(-\exp(-0.5814 - 0.013007 \cdot 8 + 1.451553)) = 0.883673$, **and**

$P^0(very\ dissatisfied, dissatistied, neutral, or\ satisfied) = 1 - \exp(-\exp(0.8883 - 0.013007 \cdot 8 + 1.451553)) = 0.999913$.

The predicted probabilities for the individual levels of the satisfaction score are:

$P^0(very\ dissatisfied) = 0.113203$, $P^0(dissatisfied) = 0.414765 - 0.113203 = 0.301562$, $P^0(neutral) = 0.883673 - 0.414765 = 0.468908$, $P^0(satisfied) = 0.999913 - 0.883673 = 0.11624$, **and** $P^0(very\ satisfied) = 1 - 0.999913 = 0.000087$.

<u>Problem 3.</u> The number of defective items produced by a machine operator during one shift is measured along with the length of work experience as a machine operator (in years), and whether it was morning, day, evening, or night shift. The data were obtained for 45 randomly chosen shifts and operators from among the pool of operators who produced at least one defective item during their shifts. Use the code and output to answer these questions:

(a) Give the name of the model that is appropriate in this case. Specify the fitted model and give estimates for <u>all</u> parameters.

In the fitted zero-truncated Poisson model, the estimated parameter is

$\hat{\lambda} = \exp(0.8369 - 0.07895 \cdot \#of\ years\ of\ experence + 0.1271 \cdot day\ shift + 0.08197 \cdot evening\ shift + 0.6999 \cdot night\ shift)$.

(b) Discuss the significance of predictors at the 5% level of significance. Give an interpretation of the estimated significant regression coefficients.

Number of years of experience and the indicator of the night shift are significant predictors. As the number of years of experience increases by one, the estimated average number of defective items produced changes by $(\exp(-0.07895) - 1) \cdot 100\% = -7.591\%$, that is, decreases by 7.591%. For the night shift, the estimated average number of defective items produced is $\exp(0.6999) \cdot 100\% = 201.36\%$ of that for the morning shift.