3. (2.5 points)

An actuary is considering using a generalized linear model to estimate the expected frequency of a recently introduced insurance product.

Given the following assumptions:

- The expected frequency for a risk is assumed to vary by state and gender.
- A log link function is used.
- A Poisson error structure is used.
- The likelihood function of a Poisson is

  $l(y;\mu)=\sum \ln f(y_i;\mu_i)=\sum -\mu_i + y_i \ln \mu_i - \ln(y_i!)$
- β_1 is the effect of gender = Male.
- β_2 is the effect of gender = Female.
- β_3 is the effect of State = State A.

Claim Frequency

	State A	State B
Male	0.0920	0.0267
Female	0.1500	0.0500

Given that $\beta_3 = 1.149$, determine the expected frequency of a male risk in State A.

CONTINUED ON NEXT PAGE

QUESTION 3

Total Point Value: 2.5 **Learning Objective: A3**

Sample Answers

Sample 1

Need Form: $Y = g^{-1}(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) + \varepsilon$

$g(x) = \log(x)$

$g^{-1}(x) = e^{\beta x}$

We have from the four observations:

$0.092 = e^{\beta_1 + \beta_3}$

$0.0267 = e^{\beta_1}$

$0.15 = e^{\beta_2 + \beta_3}$

$0.05 = e^{\beta_2}$

Loglikelihood: $\ell = \sum_{i=1}^{4} -\mu_i + y_i \ln \mu_i - \ln(y_i!)$

$$= -(e^{\beta_1+\beta_3}) + 0.092 \ln\left(e^{\beta_1+\beta_3}\right) - \ln(0.092!)$$
$$- (e^{\beta_1}) + 0.0267 \ln\left(e^{\beta_1}\right) - \ln(0.0267!)$$
$$- (e^{\beta_2+\beta_3}) + 0.15 \ln\left(e^{\beta_2+\beta_3}\right) - \ln(0.15!)$$
$$- (e^{\beta_2}) + 0.05 \ln\left(e^{\beta_3}\right) - \ln(0.05!)$$
$$= -(e^{\beta_1+\beta_3}) + 0.092(\beta_1 + \beta_3) - \ln(0.092!)$$
$$- (e^{\beta_1}) + 0.0267(\beta_1) - \ln(0.0267!)$$
$$- (e^{\beta_2+\beta_3}) + 0.15(\beta_2 + \beta_3) - \ln(0.15!)$$
$$- (e^{\beta_2}) + 0.05(\beta_2) - \ln(0.05!)$$

$$\Rightarrow \frac{d\ell}{d\beta_1} = -\left(e^{\beta_1+\beta_3}\right) + 0.092 - e^{\beta_1} + 0.0267 = 0$$
$$= -\left(e^{\beta_1+1.149}\right) + 0.092 - e^{\beta_1} + 0.0267 = 0$$
$$\Rightarrow \left(e^{\beta_1+1.149}\right) + e^{\beta_1} = 0.1187$$
$$\left(e^{1.149} \cdot e^{\beta_1}\right) + e^{\beta_1} = 0.1187$$
$$e^{\beta_1} = 0.02857$$
$$\beta_1 = -3.555$$

Expected frequency of Male state A= $e^{\beta_1+\beta_3} = e^{-3.555+1.149} = \mathbf{0.09}$

Sample 2

$$\begin{matrix} MA \\ MB \\ FA \\ FB \end{matrix} \begin{pmatrix} 0.092 \\ 0.0267 \\ 0.15 \\ 0.05 \end{pmatrix} = g^{-1}\left[\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}\right]$$

After $g^{-1}()$ => right side becomes = $\begin{pmatrix} e^{\beta_1+\beta_3} \\ e^{\beta_1} \\ e^{\beta_2+\beta_3} \\ e^{\beta_2} \end{pmatrix} = \vec{\mu}$

$$\ell(y;\mu) = \sum_{i=1}^{4} -\mu_i + y_i \, ln(\mu_i) - ln(y_i!)$$
$$= -(e^{\beta_1+\beta_3}) + 0.092(\beta_1+\beta_3) - ln(0.092!)$$
$$-(e^{\beta_1}) + 0.0267(\beta_1) - ln(0.0267!)$$
$$-(e^{\beta_2+\beta_3}) + 0.15(\beta_2+\beta_3) - ln(0.15!)$$
$$-(e^{\beta_2}) + 0.05(\beta_2) - ln(0.05!)$$

ignore all constant terms, $-\ln(y_i!)$

$$\frac{d\ell}{d\beta_1} = 0 = -e^{\beta_1+\beta_3} + 0.092 - e^{\beta_1} + 0.0267$$
$$0 = -e^{\beta_1+1.149} + 0.092 - e^{\beta_1} + 0.0267$$
$$0.1187 = e^{\beta_1}(1+e^{1.149})$$
$$\beta_1 = -3.5555$$

$$\frac{d\ell}{d\beta_2} = 0 = -e^{\beta_2+\beta_3} + 0.15 - e^{\beta_2} + 0.05$$
$$0.2 = e^{\beta_2}(1+e^{1.149})$$
$$\beta_2 = -3.0338$$

	State A	State B
M	$e^{-3.5555+1.149} = \boxed{0.09013}$	$e^{-3.5555} = 0.02857$
F	$e^{-3.0338+1.149} = 0.1519$	$e^{-3.0338} = 0.04813$

Examiners Report

This question tested the candidate's ability to formularize and solve a Generalized Linear Model. With the CAS's continued focus on advanced analytics, the process and execution of solving a GLM is something that candidates are expected to know and is included within the Learning Objective on the syllabus.

One common mistake was that candidates used the SSE instead of a loglikelihood function when plugging in the covariates. In the case of a Classical Linear Model, solving the GLM produces results that are identical to those derived when minimizing the Sum of Squared Errors for a CLM. Because this was not a CLM, the SSE could not be used to solve the problem.

Candidates generally did well identifying the X and β matrices, but struggled relating these two matrices to the Y matrix. Candidates should be aware of the relationship between these 3 matrices.

Candidates struggled to identify and apply the log link function correctly. Partial credit was given if the loglikelihood function was used but an incorrect link function was used with the covariates.

Candidates generally knew to take the partial derivative of the loglikelihood function with respect to each β and set these partial derivatives to 0 in order to solve for the β's. Candidates did not need to solve for β_2 to receive full credit for the problem, and points were not taken off for solving for this parameter.

Common mistakes included:

- Not identifying and applying the log link function correctly
  - Incorrectly stating the relationship between the 3 matrices (multiplying the X and β matrices together, without using the correct $g^{-1}(x)$ function, and setting equal to the Y matrix)
  - Using the logit link function instead of the log link function
  - Not recognizing that μ needed to equal $e^{\Sigma X\beta}$. Candidates frequently set μ equal to $\sum X\beta$ in the loglikelihood function
- Changing the β's given in the problem, including:
  - Adding β_0, or some intercept term
  - Removing a β that was given. There were 4 factors in the problem and 3 covariates were given. These did not need to be adjusted, since there was no aliasing in the problem
- Using the SSE instead of the given loglikelihood function
- Not stating that the partial derivatives with respect to each β need to be set equal to 0
- Calculation errors with derivatives, exponentials and ln()

  - For example, $e^{\beta_1+1.149} + e^{\beta_1} \neq e^{2\beta_1+1.149}$
- Using the incorrect formula for the expected frequency for a Male in State A