7. (2.5 points)

An actuary is planning to add a credit-based insurance score to a model that estimates the probability of a policy having a claim. The actuary has decided to offset all of the current model variables before fitting the new variable.

Given the following:

- The current model (without the insurance score variable) is a logit-link binomial GLM (logistic regression).
- The logit link function is defined as $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$
- The insurance score is a continuous variable having a value between 1 and 100.
- The current fitted values and insurance score for three policies as well as regression results from the fit of the insurance score variable are given below:

Policy Number	Fitted Probability Without Insurance Score	Insurance Score
1	1.3%	78
2	20.3%	92
3	2.5%	35

Variable	Parameter Estimate
Intercept	1.250
Insurance Score	-0.020

a. (0.5 point)

Calculate the offset term to be used in the regression for each of the three policies above.

b. (0.75 point)

Calculate the revised fitted probability of having a claim for each of the three policies above.

c. (0.5 point)

Identify the range of:

i. the logit function
ii. the logistic function

d. (0.25 point)

Briefly explain why logistic regression is often used to model probabilities.

e. (0.5 point)

Identify and briefly describe one situation in which the use of an offset is preferable to (re)fitting all variables.

CONTINUED ON NEXT PAGE

SAMPLE ANSWERS AND EXAMINER'S REPORT

QUESTION 7	
TOTAL POINT VALUE: 2.5	**LEARNING OBJECTIVE(S): A3**
SAMPLE ANSWERS	
Part a: 0.5 point	
Policy 1: ln(0.013/(1-0.013) = -4.330 Policy 2: ln(0.203/(1-0.203) = -1.368 Policy 3: ln(0.025/(1-0.025) = -3.664	
Part b: 0.75 point	
Policy 1: $\ln(\mu/(1-\mu)) = -4.330+1.25-.02x78 = -4.640$ $\mu = 1.0\%$ Policy 2: $\ln(\mu/(1-\mu)) = -1.368+1.25-.02x92 = -1.958$ $\mu = 12.4\%$ Policy 3: $\ln(\mu/(1-\mu)) = -3.664+1.25-.02x35 = -3.114$ $\mu = 4.3\%$	
Part c: 0.5 point	
Logit Function: $\ln(\mu/(1-\mu))$ Value range $(-\infty,\infty)$ Logistic Function: $1/(1+ e^{-\mu})$ Value range (0,1)	
Part d: 0.25 point	
Because the logistic function can take any value from $(-\infty,\infty)$ and map it to a value between 0 and 1. This is also the range for probabilities, so it is an intuitive fit.	
Part e: 0.5 point	
Sample 1 Coverage-related options on a policy. This is because there are often counterintuitive relativities due to selection effects, so best to calculate in separate model and include as an offset. *Sample 2* Offset is useful for deductible, which is better estimated outside GLMs (e.g. LER analysis), since GLM often produces counterintuitive results due to effect of selection and correlation with variables outside model. *Sample 3* Territory rating is impractical to use in a GLM since there are hundreds or even thousands of territories with no easy way to group them without losing signal. However, territory differences are significant so it's important that the rating plan be offset for territory rates. Thus it's best to include territory factors as an offset in GLM. *Sample 4* If you're creating a model on renewal business after having already made a model for new business only, you would likely use an offset for many of the variables. This would ensure consistency between the sets of business that you do not expect to change over time.	

Sample 5

When including the effect of a coverage limit in a pure premium model. Limits may be correlated with other covariates not being accounted for in the model and this might lead to inconsistent ILFs based on model results, so it's better to do loss elimination analysis outside of the modeling process and include the effect of a coverage limit as an offset.

Sample 6

When introducing additional variables but do not want to change existing ones due to constraints like rate filing approval, IT system constraints, etc.

EXAMINER'S REPORT

Candidates were expected to know how to calculate an offset given the results of a prior model and then to use that offset in the new model with the addition of a new variable to calculate the new target variable (probability). They were also expected to know when it is appropriate to use an offset in a GLM. Candidates were expected to know the relationship between the logit function and logistic function and why the logistic function is appropriate for modeling probabilities.

Common mistakes included:

- Interchanging the logit and the logistic function
- Using the wrong link function
- Not tying all the pieces of the GLM (link function and linear predictor components such as intercept, insurance score and offset, etc) together

Part a

Candidates were expected to calculate the offset term to be used in the regression for each of the three policies.

Common mistakes included:

- Including the intercept or insurance score
- Using the incorrect link function

Part b

Candidates were expected to calculate the revised probability of a claim, using the model with the insurance score, assuming the old model as an offset.

Common mistakes included:

- Taking the natural log of the credit score
- Not including the insurance score (if it wasn't also erroneously included in part a)
- Not including the intercept (if it wasn't also erroneously included in part a)
- Not including the offset calculated in part a
- Not using the correct formula for μ

Part c

Candidates were expected to identify the range of the logit and logistic functions.

A common mistake was identifying the input range as opposed to the output range.

Part d

Candidates were expected to explain why logistic regression is used to model probabilities.

Common mistakes included:

- Just mentioning that it's in the probability range without giving what that range was
- Mentioning that logistic regression uses a binomial response (Y/N) without tying it to the range of the logistic function
- Mentioning the logit function/odds ratio, without tying it to the range of the logistic function

Part e

Candidates were expected to identify a situation in which an offset would be preferable as well as provide a description for that situation.

A common mistake was not fully describing the situation or not fully providing an example.