6. (2.5 points)

An actuary is comparing the output of two generalized linear models to develop a new rating plan for personal auto. Model statistics are shown below:

	Saturated Model	Model A	Model B
Log-Likelihood	-1,000	-1,500	-1,465
Estimated Dispersion Parameter	1.75	1.75	1.75

- Model A is a nested model of Model B, where Model B has an additional variable for driver age.
- Driver age is fit using a second-order polynomial.
- The critical value to be used from the F-distribution is 19.5.

a. (2 points)

Using two statistical tests, recommend whether or not driver age should be included in the rating plan.

b. (0.5 point)

Describe why the deviance statistic alone should not be used to assess model fit.

CONTINUED ON NEXT PAGE

QUESTION 6	
TOTAL POINT VALUE: 2.5	**LEARNING OBJECTIVE(S): A4**
SAMPLE ANSWERS	

Part a: 2 points

Sample 1

F-test

$$F = \frac{D_S - D_B}{\Phi_S \bullet p_{added}} = \frac{1000 - 930}{1.75 \bullet 2} = 20 > 19.5 \Rightarrow reject\ H_O\ \&\ use\ Model\ B$$

D = $2(\ell\ell_{\text{sat}} - \ell\ell_{\text{model}})$

$D_S = 2(\text{-}1000 - (\text{-}1500)) = 1000$

$D_B = 2(\text{-}1000 - (\text{-}1465)) = 930$

=>second degree polynomial; adding 2 params

AIC = $-2\ell\ell + 2p$

AIC_S = -2(-1500) + $2p_S$ = 3000 + $2p_S$

AIC_B = -2(-1465) + 2($p_S + 2$) = $\underline{2934 + 2p_S}$

66 > 0 -> model B better (smaller AIC)

Based on both F test & AIC, model B is better

=> Include age in rating plan

Sample 2

AIC = -2LL + 2p

AIC_A = -2(-1500) + 2p = 3000 +2p

AIC_B = -2(-1465) + 2(p+2) = 2930 +2p + 2

AIC_B = 2934 +2p < 3000 + 2p -> Model B is better based on AIC

BIC = -2LL + p•ln(n)

BIC_A = -2(-1500) + p•ln(n) = 3000 + p•ln(n)

BIC_B = -2(-1465) + (p+2)•ln(n)= 2930 + (p+2)•ln(n)

3000 + ~~p•ln(n)~~ = 2930 + ~~p•ln(n)~~ + 2•ln(n)

ln(n) = 35 n = e^{35}

Assume the number of observations (n) is less than e^{35}, so $BIC_B < BIC_A$ so Model B is better based on BIC as well

∴ I recommend including age since Model B performs better on both tests.

Sample 3

F-Test:

F statistic $= \frac{2(\ell\ell_B - \ell\ell_A)}{\Delta parameter \bullet \Phi_{Small}}$

$= \frac{2(-1465+1500)}{2 \bullet 1.75} = 20 > 19.5$

B is better than A

➔ Should include driver age

AIC: Assume # parameters in A is 20, since not given
AIC (A) = -2(-1500) + 2•20 = 3040
AIC (B) = -2(-1465) + 2•(20+2) = 2974
Since AIC is lower for B, B is superior

Thus I recommend using Model B as it performs better using the F test and AIC criteria.

Part b: 0.5 point

Sample 1
Because adding more variables into the model will always reduce the deviance statistic which will cause the model to be overfit.

Sample 2
Deviance isn't useful as adding more variables always decreases the deviance. Using AIC or BIC is more appropriate, as they penalize for adding new parameters.

EXAMINER'S REPORT

- Candidates were expected to know the formulas for F-test, AIC test, and/or BIC test and to be able to conclude whether the results from the test indicated inclusion or exclusion of the new variable.
- Candidates were expected to know that deviance and log-likelihood are inappropriate measures for comparing model structure in the situation given.
- Candidates were expected to be able to determine the number of parameters added to the model based on the given model form.

Part a

Candidates were expected to calculate two test statistics from any two of the F-test, AIC test, and BIC test, and to include a final conclusion that combined the results of both tests.

- Every test required that the candidate demonstrate that he/she knew how many parameters were added to the model.
- Only partial credit was given for using deviance or log-likelihood to compare models. These are not adequate for comparing nested models, which is the topic of part b. The conclusion could be made without calculating deviance as we know what it will be.
- Candidates should note that, as of this sitting, there is an ambiguity in the source text: the source paper does not adequately distinguish between deviance and scaled deviance. Therefore, the F-test statistic in this solution, while consistent with the source text, is technically incorrect. Using the correct methodology, the F-statistic would be calculated as follows:

  Model A Scaled Deviance = (2)(-1,000 + 1,500) = 1,000
  Model B Scaled Deviance = (2)(-1,000 + 1,465) = 930
  Model A Deviance = 1,000 * 1.75 = 1,750
  Model B Deviance = 930 * 1.75 = 1,627.5
  F stat = (1,750-1,627.5) / (2 * 1.75) = 35

Had a candidate performed the above calculation, they would have received full credit. The source text is currently being revised, and future candidates should make note of this when using this report as a study resource.

Common mistakes included:

- Failing to make an overall conclusion that combined the results of both tests
- Performing only one test
- Using 1 or 3 rather than 2 as the additional number of parameters
- Incorrect calculation of the deviance used in the numerator of the F-test. Specifically, candidates often forgot to multiply by two resulting in an F statistic that was half as large as it should have been and causing them to make the incorrect conclusion
- Assuming that the AIC or BIC could not be done without knowing the number of parameters. For AIC the number of parameters cancels out when comparing the models and becomes unimportant. For BIC, credit was given for either assuming a number of observations or stating at what cutoff the number of observations would change the conclusion. Credit was also given if candidates made a statement about the assumed number of parameters and/or the assumed number of observations.
- Using deviance in the AIC/BIC formula rather than log-likelihood or using log-likelihood rather than deviance in the calculation of the F-test.
- Making the wrong conclusion on a particular test even with the correct calculations.

Part b

Candidates were expected to know that deviance decreases or that log-likelihood increases with the addition of variables. Credit was not given if candidates simply said it improves.

Candidates were expected to know that using deviance alone would lead to over-fitting. Credit was also given to recognizing fitting to noise, as well as statements about penalizing for adding additional parameters.

Common mistakes included:

- Giving some of the limitations of deviance such as needing to have the same underlying dataset with the same distribution. This limitation is not restricted to deviance alone and it addresses situations where deviance or any test based on deviance is not appropriate at all for comparison rather than the question of why more than one test should be considered in the situations where deviance and tests based on deviance are appropriate to use for model comparison.