1. **Poisson Regression** Please refer to Dobson Exercise 9.2 for a description of the data table.
   Data File Table 9.13-Claims.xlsx attached.
   (a) EDA - Please calculate the rate of claims y/n for each category. Produce scatterplots of the claims rate by AGE, and the claims rate by CAR. Produce a side-by-side boxplot of claims rate by DIST. Describe your visual impression of the main effects of these factors.
   (b) Treat AGE, CAR and DIST as categorical and use Poisson regression to choose between models with and without any interactions. (Equivalently, test the hypothesis that coefficients of interaction terms are all zero.)
   (c) Fit the model without interactions, by treating AGE and CAR as continuous variables and DIST as categorical.
      i Please specify the model and provide coefficient estimates
      ii Please manually calculate the chi-squared goodness of fit statistic (X^2)
      iii Please manually calculate the deviance statistic (or at least check if you agree with the deviance statistic calculated by R)

2. Consider a $2 \times K$ contingency table (1) in which the column total $y_{\cdot k}$ are fixed for $k = 1, 2, \ldots, K$. (Adapted from Dobson exercise 9.6.)

	1	...	k	...	K
Success	y_{11}	...	y_{1k}	...	y_{1K}
Failure	y_{21}	...	y_{2k}	...	y_{2K}
Total	$y_{\cdot 1}$	...	$y_{\cdot k}$	...	$y_{\cdot K}$

Table 1: Contingency table with 2 rows and K columns

   (a) Show that the product multiracial distribution for this table reduces to

$$f(z_1, \ldots, z_K | n_1, \ldots, n_K) = \prod_{k=1}^{K} \binom{n_k}{z_k} \pi_k^{z_k} (1 - \pi_k)^{n_k - z_k}$$

   where $n_k = y_{\cdot k}$, $z_k = y_{1k}$, $n_k - z_k = y_{2k}$, $\pi_k = \theta_{1k}$ and $1 - \pi_k = \theta_{2k}$ for $k = 1, 2, \ldots, K$. This is the **Product binomial distribution**.
   (b) Show that the log-linear model with

$$\eta_{1k} = \log \mathrm{E}[Z_k] = \mathbf{x}_{1k}^T \beta$$

   and

$$\eta_{2k} = \log \mathrm{E}[n_k - Z_k] = \mathbf{x}_{2k}^T \beta$$

   is equivalent to the logistic model

$$\log\left(\frac{\pi_k}{1 - \pi_k}\right) = \mathbf{x}_k^T \beta$$

   where $\mathbf{x}_k = \mathbf{x}_{1k} - \mathbf{x}_{2k}$, $k = 1, 2, \ldots, K$.

(c) Base on (b), analyze the case-control study data on aspirin use and ulcers, Table 9.7 Gastric and duodenal ulcers and aspirin use Dobson page 174, using logistic regression and compare the results with those obtained using log-linear models.

3. **(Given the predictions, deduce the parameters)** A GLM was used to estimate the expected losses per customer across gender and territory. The following information is provided:

   - The link function selected is log
   - Q is the base level for *Territory*
   - Make is the base level for *Gender*
   - Interaction terms included in the model

   The GLM produced the following predicted values for expected loss per customer:

Gender	**Territory**	
	Q	R
Male	148	545
Female	446	4024

Table 2: Predictions for Gender & Territory Model For Question 3

   Please Calculate the estimated beta for the interaction of Territory R and Female.

   hint

   - write the log link model $\log \mu = \beta_0 + \beta_1 x_1 + + \beta_2 x_2 + \beta_3 x_1 x_2$, where x_1, x_2 is gender and territory respectively
   - write the predicted value $\hat{\mu}$ in term of the model above.

4. **Chi-square goodness of fit**

   You are given the following claim counts:

Number of Claims	**Number of Policies**
0	2050
1	450
2	80
3	20
4 or more	0
Total	2600

Table 3: Claim Counts

   Assume that claim counts follow a Poisson distribution, and the Poisson parameter is estimated using maximum likelihood.

(a) Calculate the sample mean.

(b) Calculate the chi-square statistic.

5. You are given the following information for a fitted GLM:

Response Variable	Claim Size	
Response Distribution	Gamma	
Link	Log	
Scale parameter	$\alpha = 1$	
Parameter	df	$\hat{\beta}$
Intercept	1	2.100
Zone	4	
1	1	7.678
2	1	4.227
3	1	1.336
4	0	0.000
5	1	1.734
Vehicle Class	6	
Convertible	1	1.200
Coupe	1	1.300
Sedan	0	0.000
Truck	1	1.406
Minivan	1	1.875
Stationwagon	1	2.000
Utility	1	2.500
Drive Age	2	
Under 30	2.000	
30-49	0.000	
50 and above	1.800	

$\text{Gamma}(y; \alpha, \beta) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{y}{\beta}\right)^{\alpha-1} \exp\left(-y/\beta\right)$
where, $y, \alpha, \beta \in \mathcal{R}_+$

(a) Calculate the predicted claim size for an observation for Zone 3, with Vehicle Class Truck and Drive Age 55.

(b) Calculate Variance of a claim size for an observation for zone 4, with Vehicle Class Sedan and Drive Age 35.