Problem 1

The following table gives data from a retrospective study between esophageal cancer and daily alcohol consumption adjusted for age.

Age	Case		Control	
	Daily Alcohol Consumption		Daily Alcohol Consumption	
	0-79 g	80+ g	0-79 g	80+ g
25-34	0	1	106	9
35-44	5	4	164	26
45-54	21	25	138	29
55-64	34	42	139	27
65-74	36	19	88	18
75+	8	5	31	0

1. Fit a prospective model to the data to study the relation between alcohol consumption, age, and disease (model age as a continuous variable taking values 25, 35, 45, 55, 65, and 75). Interpret the result.

2. (Optional; required for PhD) Let Ψ_j be the odds ratio relating alcohol consumption and disease in the jth age group ($j = 1, ..., 6$). Assume different age groups have different odds. Compare the following two models:
   $M_0 : \Psi_j = 1$ for all j;

   $M_1 : \Psi_j = \Psi$ (where Ψ is an unknown constant);

   (Hint: First write out the models and check if they are nested; if so, use deviance analysis to compare the two models.)

Problem 2

The following table provides data from a study of the germination of two species of Orobanche seeds. The seeds were grown on 1/125 dilutions of two different root extract media (cucumber or bean) in a 2x2 factorial layout with replicates. The data (y_i/m_i) consist of the number of seeds, m_i, and the number germinating, y_i, for each batch. Interest focuses on the possible differences in germination rates for the two types of seed and root extract.

O. aegyptiaca 75		O. aegyptiaca 73	
Bean	Cucumber	Bean	Cucumber
10/39	5/6	8/16	3/12
23/62	53/74	10/30	22/41
23/81	55/72	8/28	15/30
26/51	32/51	23/45	32/51
17/39	46/79	0/4	3/7
	10/13		

1. Fit a logistic regression model to study the relation between germination rates and different types of seed and root extract. Interpret the result.

2. Is there over dispersion? If so, what is the estimate of dispersion parameter? Update your model and reinterpret the result.

3. What is a plausible cause of the over dispersion?