Homework I (2022)

Please solve the following problems and then submit the pdf copy of them.

1. For c categories, it assumes counts $(Y_1, Y_2, \ldots, Y_c)$ are independent Poisson(μ_j) for $j = 1, \ldots, c$, then show that given $\sum_{j=1}^{c} Y_j = n$, conditional distribution of $(Y_1, Y_2, \ldots, Y_c)$ is multinomial distribution with $\pi_j = \frac{\mu_j}{\sum_{k=1}^{c} \mu_k}$ for $j = 1, \ldots, c$.

2. In class, we learn multivariate delta method to calculate the variance of estimate. Using the result, show that the standard error of $\log \hat{\theta}$ is $\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$ in 2×2 contingency table.

3. For diagnostic testing, let X =true status (1= disease, 2=no disease) and Y =diagnosis (1=positive, 2=negative). Let $\pi_1 = P(Y = 1|X = 1)$ and $\pi_2 = P(Y = 1|X = 2)$. Let γ denote the probability that a subject has the disease.

   (a) Given that the diagnosis is positive, use Bayes' Theorem to show that the probability a subject truly has the disease is

   $$P(X = 1|Y = 1) = \pi_1\gamma / [\pi_1\gamma + \pi_2(1 - \gamma)].$$

   (b) For mammograms for detecting breast cancer, suppose $\gamma = 0.01$, sensitivity= $\pi_1 = 0.86$, and specificity $1 - \pi_2 = 0.88$. Find the positive predictive value.

   (c) To better understand the answer in (b), find the joint probabilities for the 2×2 cross-classification of X and Y. Discuss their relative sizes in the two cells that refer to a positive test result.

4. The following table cross-classifies votes in the 2008 and 2012 US Presidential elections. Estimate and find a 95% confidence interval for the population odds ratio. Interpret.

	Vote in 2012	
Vote in 2008	Obama	Romney
Obama	802	53
McCain	34	494

Source: 2014 General Social Survey.