STAT511 HW#10

Reading: Read Chapter 10 of Ott & Longnecker
See Canvas Calendar for Due Date.

30 points total, **2** points per problem part unless otherwise noted.

1. A case-control study in Berlin, reported by Kohlmeier, Arminger, Bartolomeycik, Bellach, Rehm and Thamm (1992) and by Hand et al. (1994) asked 239 lung cancer patients and 429 healthy controls (matched to the cases by age and sex) whether or not they had kept a pet bird during adulthood. The data is summarized below:

	Cancer Patients	Healthy Controls	Total
Bird	98	101	199
No Bird	141	328	469
Total	239	429	668

   A. Estimate the odds ratio (of "Cancer" versus "Control") for the "Bird" versus "No Bird" groups. Does the Bird or No Bird group have higher odds of lung cancer?
   B. Give a 95% confidence interval for the odds ratio. Based on this interval, can you conclude that there is a relationship between bird ownership and lung cancer? NOTE: Use method="wald".
   C. Now run the chi-squared test for this data. Give the p-value and conclusion.

2. Bacillus Calmette-Guerin (BCG) is a vaccine for preventing tuberculosis. For this question we will examine data from 3 studies (Vandiviere et al 1973, TPT Madras 1980, Coetzee & Berjak 1968). The data is summarized below.

Study	Trt Status	TBpos	TBneg
1	Trt	8	2537
1	Ctrl	10	619
2	Trt	505	87886
2	Ctrl	499	87892
3	Trt	29	7470
3	Ctrl	45	7232

   A. Calculate the odds ratio for each study separately. (**3 pts**)
   B. Use the Breslow-Day test to test for equality of odds ratios across the 3 studies. State your p-value and conclusion. Can we conclude that the odds ratios are different across the 3 studies?
   C. Typically, we may not want to pool information across studies if the estimated odds ratios were not equal. However, for this question run the CMH test anyway and present the common estimate of the odds ratio, p-value and conclusion (**3 pts**)

A note about the BCG vaccine from Wikipedia:
The most controversial aspect of BCG is the variable efficacy found in different clinical trials that appears to depend on geography. Trials conducted in the UK have consistently shown a protective effect of 60 to 80%, but those conducted elsewhere have shown no protective effect, and efficacy appears to fall the closer one gets to the equator.

3. The table below presents count data on the number of *Larrea divaricata* plants found in each of 48 sampling quadrants, as reported in the paper, "Some sampling characteristics of plants and arthropods of the Arizona desert," (Ecology, 1962, 567-571). Let Y = number of plants in a quadrant.

Observed Counts							
Number of Plants (Y)	0	1	2	3	4	5	6
Frequency	9	9	10	14	2	2	2

   A. Find the sample mean (point estimate of the population mean), $\hat{\mu}$.
   B. To investigate the behavior of *Larrea divaricata* plants, researchers have decided to test whether the number of plants (Y) follows the Poisson distribution. Since the cells with $Y \geq 4$ have small frequencies, they have decided to combine all the cells with $Y \geq 4$.

Observed Counts					
Number of Plants (Y)	0	1	2	3	≥4
Frequency	9	9	10	14	6

   Using the table right above, Calculate the GOF test statistic, p-value, and give a conclusion using $\alpha = 0.05$. (**4 pts**)

4. The data "PoissonData.csv" was generated from the Poisson distribution (using the rpoiss() function).
   A. Calculate the sample mean, sample standard deviation. Also construct a histogram and qqplot of the data. (**4 pts**)
      NOTE: Because the data comes from the Poisson distribution, you should find that the mean and the sample variance (s^2) are close. However, you should also find from the histogram and qqplot that the data looks approximately normal.
   B. Give a standard t-based 95% confidence interval for μ.
   C. Following the example on CH10 Slide 99, construct a 95% confidence interval for μ based on the normal approximation to the Poisson distribution. (**4 pts**)

   NOTE: The CIs from parts B and C should be similar.