STAT 513 hw 9

1. Suppose a study of the efficacy of a treatment results in the following table of outcomes:

	Successes	Failures	Total
Treatment	20	5	25
Control	10	16	26
Total	30	21	51

   (a) State the null and alternate hypotheses which are likely of interest to the researchers.
   (b) i. Compute the test statistic for the likelihood ratio test of your hypotheses from part (a).
       ii. Give the p-value.
   (c) i. Compute the test statistic for Pearson's chi-squared test of your hypotheses from part (a).
       ii. Give the p-value.
   (d) Give the p-value of Fisher's exact test of your hypotheses in part (a).

2. The following knee injury data in women collegiate rugby players is taken from [2]. It is of interest to know whether the types of injuries a player experiences are associated with the position (Forward, Back) she plays.

	Meniscal Tear	MCL Tear	ACL Tear	Other
Forward	13	14	7	4
Back	12	9	14	3

   (a) i. Compute the test statistic for the likelihood ratio test of association.
       ii. Give the p-value (make sure you choose the right degrees of freedom!).
   (b) i. Compute the test statistic for Pearson's chi-squared test of association.
       ii. Give the p-value.

3. Consider the following data taken from [1], which result from looking through a microscope at samples of milk film and counting the number of bacterial colonies within the field of vision. A total of 400 observations were gathered and the number of bacterial colonies was recorded for each of them:

# Bacterial Colonies	# Microscopic fields
0	56
1	104
2	80
3	62
4	42
5	27
6	9
7	9
8	5
9	3
10	2
19	1

(a) Let X_i represent the number of bacterial colonies in microscopic field i, with $i = 1, \ldots, 400$. Assume for a moment that the number of bacterial colonies in a microscopic field follows the $\text{Poisson}(\lambda)$ distribution for some value of λ. Compute the maximum likelihood estimator of λ based on the above data.

(b) If the bacterial colony counts truly followed a Poisson distribution with mean equal to $\hat{\lambda}$, where $\hat{\lambda}$ is the maximum likelihood estimator of λ computed in part (a), what would be the expected # of microscopic fields corresponding to each number of bacterial counts? That is, in how many microscopic fields out of 400 would we expect to see 0 bacterial colonies, 1 bacterial colony, and so on? Make a table like the table above, but with the numbers in the right-hand column replaced by the expected numbers of microscopic fields. *Hint: the first one is* `dpois(0,lambda.hat)*400`.

(c) From looking at these numbers, do you believe the # of bacterial colonies follows a Poisson distribution? Explain your answer.

(d) Pearson's chi-squared test is often used to test for what is called the "goodness-of-fit" of a probability distribution to some observed data, as this test statistic provides a useful way to compare observed counts to expected counts. Compute the test statistic of Pearson's chi-squared test on these data, which is given by

$$\sum_{i=1}^{12} (O_i - E_i)^2 / E_i,$$

where $O_1, \ldots, O_{12}$ are observed numbers of microscopic fields and $E_1, \ldots, E_{12}$ are the expected numbers of microscopic fields according to the $\text{Poisson}(2.44)$ distribution. *Hint: You get a crazy-huge number.*

(e) It was naïve of us to compute Pearson's test statistic in the previous part, because some of the expected counts are very small, almost equal to zero; recall that we require expected counts to be greater than or equal to 5 in order to use Pearson's chi-squared test. What can we do? Let's collapse the last few rows of the tables by summing together the rows for which the # of bacterial colonies is greater than or equal to 6, so that we have the following:

# Bacterial Colonies	# Microscopic fields	$\mathbb{E}$[# Microscopic fields]
0	56	34.86
1	104	85.07
2	80	103.78
3	62	84.41
4	42	51.49
5	27	25.13
≥ 6	29	15.23

Recompute the test statistic for Pearson's chi-squared test, this time with

$$\sum_{i=1}^{7} (O_i - E_i)^2 / E_i.$$

(f) Under the null hypothesis, the test statistic converges in distribution to a random variable with the χ^2_6 distribution, since we are considering a table with 7 cells in a column, and $7 - 1 = 6$. Use this information to compute a p-value for the test statistic computed in part (e). Use the p-value to make a conclusion about whether or not the # of bacterial colonies follows a Poisson distribution.

References

[1] Chester Ittner Bliss and Ronald A Fisher. Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2):176–200, 1953.

[2] Andrew S. Levy, Merrick J. Wetzler, Marie Lewars, and William Laughlin. Knee injuries in women collegiate rugby players. *The American journal of sports medicine*, 25(3):360–362, 1997.