StatsI — Exercise 8

1. Suppose that $X_1, ..., X_n$ and $Y_1, ..., Y_n$ are two independent random samples from two exponential distributions with mean μ_1 and μ_2 respectively. Find the likelihood ratio test for $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Specify the asymptotic distribution of the test statistic under H_0.

2. A survey of the use a particular product was conducted in four areas, and a random sample of 200 potential users was interviewed in each area. In area i, for $i = 1, 2, 3, 4$, X_i of the 200 said that they used the product. Construct a likelihood ratio test to test whether the proportion of the population using the product is the same in each of the four areas. Carry out the test at 5% level for the case $X_1 = 76, X_2 = 53, X_3 = 59$ and $X_4 = 48$.

3. In 1861, 10 essays appeared in the *New Orleans Daily Crescent*. They were signed "Quintus Curtius Snodgrass" and some people suspected they were actually written by Mark Twain. To investigate this, we will consider the proportion of three letter words found in an author's work. From 8 Twain's essays, the proportions are:

   0.225, 0.262, 0.217, 0.240, 0.230, 0.229, 0.235, 0.217.

   From 10 Snodgrass essays, the proportions are:

   0.209, 0.205, 0.196, 0.210, 0.202, 0.207, 0.224, 0.223, 0.220, 0.201.

   (a) Perform a Wald test for equality of the means. Report the p-value and a 95% confidence interval for the difference of means. What do you conclude?

   (b) Now use a permutation test to avoid the use large sample methods. What is your conclusion?

4. An air-traffic controller claims that the number of radio messages received per minute is a random variable having a Poisson distribution with the mean $\lambda = 1.5$. The numbers of radio messages received in 200 one-minute intervals are recorded in the table below. Does this data set support the claim made by the air-traffic controller? If not, can you identify another Poisson distribution which provides a better goodness-of-fit?

No. of radio message	0	1	2	3	4	5
Frequency	70	57	45	21	5	2

5. The table below summarized the fate of the passengers and the crew when the *Titanic* sank on Monday, 15 April 1912. Test the hypothesis of independence between the row variable and the column variable in the table, and interpret your findings.

		Gender/Age Category			
		Men	Women	Boys	Girls
Fate	Survived	332	318	29	27
	Died	1360	104	35	18

6. For the 2×2 contingency table below, show that the absolute difference between the observed frequency and the expected frequency under the independence hypothesis is $|ad - bc|/n$ for all the four cells, where $n = a + b + c + d$.

$$\begin{array}{cc|c} a & b & a+b \\ c & d & c+d \\ \hline a+c & b+d & n \end{array}$$

Furthermore, the goodness-of-fit test statistic for the independence is

$$T = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

7. The following data pertain to the quality of shipments received by a firm from two different suppliers.

	No. of shipments rejected	No. of shipments accepted	
Supplier A	30	70	100
Supplier B	25	175	200
	55	245	300

(a) Test whether the quality of shipments is independent of the particular supplier.

(b) It is claimed that the probability of a rejected shipment from Supplier A is the same as that from Supplier B. Perform an appropriate test for this hypothesis.