Topics to be Covered this week

STA 138
Fall Quarter, 2018

Monday, October 8 Pearson's chi-square and likelihood ratio tests (Handout 4), Two-way contingency tables (chap 2.1 in the text and Handout 5).

Wednesday, October 10 Two-way contingency tables (chaps 2.1-2.3 in the text and Handout 5).

Friday, October 12 Tests for independence, Exact Inference for small samples (chaps 2.4 and 2.6 in the text, and Handouts 5 and 6).

Homework 2: Due on Monday, October 15.

You may form a group of 3 students registered in this course and submit one completed homework for the group. The front page should display only the names of the students in the group. The actual work should start from the second page.

1. A demographer claims that 15% of a certain racial group are left handed. In a random sample of 150 from this race 13 turns out to be left handed.
(a) Write down the null and the alternative hypotheses.
(b) Calculate the likelihood ratio statistic for this testing problem.
(c) Carry out the test using the likelihood ratio statistic at level $\alpha = 0.05$, and state your conclusion. Find the p-value of your test?

2. A company has developed a new inexpensive diagnostic method for early screening for a certain disease. In order to examine reliability of this method, a physician takes a random sample of 300 patients with the disease. All the sample individuals are given the diagnostic test, and counts are given in the table below.

	Diagnosis		
	Positive	Negative	Inconclusive
Count	233	60	7

On its website the company states that for those with the disease, the overall results of their diagnostic method are: 70% 'Positive' , 25% 'Negative' and 5% 'Inconclusive'.
(a) It is desired to carry out a test to verify the claim of the company. Write down the null and the alternative hypotheses. Also obtain the expected counts under the null hypothesis.
(b) Calculate the value of Pearson's chi-square statistic for this testing problem.
(c) Find the p-value of the test and state your conclusion at a level $\alpha = 0.05$

3. In a certain country 30% are seniors, and 40% of the seniors are vitamin D deficient, and among the non-seniors, 20% are vitamin D deficient. If a person is chosen randomly, let X be the age (1=senior, 2=not senior), and let Y be the vitamin D status (1=deficient, 2=not deficient).

(a) Let $\pi_{ij} = P(X = i, Y = j)$. Obtain the table for the joint probabilities $\{\pi_{ij}\}$.
(b) Find the odds for vitamin D deficiency among seniors and among non-seniors. Also obtain the odds ratio for vitamin deficiency for seniors to non-seniors. Interpret this odds ratio.
(c) Are age and vitamin D status independent? Explain your answer.

4. Is vitamin C effective in preventing common cold. In a Canadian study, 818 volunteers were randomly split into two groups during the cold season: vitamin C group (received 1000mg of vitamin C per day), placebo (inert pills). At the end of of the cold season, a physician determined if each sample subject had a cold or not. The counts are given below.

	Outcome	
	Cold	No Cold
Placebo	335	76
Vitamin C	302	105

Assume that the 818 volunteers form a random sample of the population in Canada.
(a) Let π_1 the population proportion catching cold among those who do not take vitamin C. Let π_2 be the corresponding population proportion among those who take vitamin C. Obtain the MLE for π_1 and π_2, odds for catching cold for placebo and vitamin C groups, estimate the odds ratio of cold for the placebo group to the vitamin C group. Interpret the odds ratio.
(b) Obtain a 95% confidence interval for $\pi_1 - \pi_2$, and interpret it
(c) Obtain a 95% confidence intervals for the odds ratio. Does the result indicate if catching cold is independent of vitamin C status? Explain.

5. A gambling game involves rolling three dice, and the number of sixes rolled is proportional to your winnings. A gambler plays the game 202 times, and the table below lists the distribution of sixes. The goal is to check if the gambler distribution matches what we would expect to see with fair dice.

# of sixes	0	1	2	3
Frequency	96	70	30	6

(a) Write down the null hypothesis for the relevant test.
(b) Calculate the value of the likelihood ratio statistic.
(c) Find the p-value of your test. Do you believe the dice the gambler was using were rigged, or fair? Explain.

6. In order to investigate how the educational levels of adults are related to their race, two surveys were done in a major metropolitan area.

In survey 1, 225 adults were randomly taken, and their highest educational level (1=none, 2=high school, 3=college) as well as their race (1=black, 2=white) were recorded.

In survey 2, 100 blacks were randomly selected and independently, 100 whites were selected. Educational level of each of the sample subjects was recorded.

Denote by π_{ij} be the proportion of individuals who are of the i^{th} race with j^{th} educational level in the metropolitan area.

The data are given below.

Survey 1

	Educational Level			
	None	High School	College	Row Total
Race				
Black	25	62	37	124
White	20	40	41	101
Column Total	45	102	78	225

Survey 2

	Educational Level			
	None	High School	College	Row Total
Race				
Black	21	49	30	100
White	22	29	49	100
Column Total	43	78	79	225

The questions below are conceptual in nature, and numerical calculations are not necessary. Please provide adequate explanation for each answer.

(a) For each of the two surveys, describe if the distribution of the counts $\{n_{ij}\}$ is jointly multinomial or independent multinomials. Write down the parameters of the distributions.

(b) The Welfare Department of the state is interested in estimating π_{ij}'s, π_{i+}'s, π_{+j}'s. Can you estimate these from the data in Survey 1? Survey 2?

(c) The state is interested in obtaining estimates of the proportion of None, High School and College graduates among blacks as well as among whites. Can these be estimated from the data in Survey 1 and/or Survey 2?

(d) The state would like to have estimates of the percentage of blacks and whites among those who are college graduates. Can you obtain these estimates from Survey 1 data? How about Survey 2 data?

(e) It is of interest to test if educational level is independent of race. Can this done using the data from Survey 1? How about Survey 2 data?