Topics to be covered this week.

STA 138

Monday, Oct 15 Tests for independence in a two-way table (chap 2.4 in the text, and Handout 5)

Wednesday, Oct 17 Test of independence in a two-way table and partitioning of chi-square (chap 2.4 in the text and Handout 6), Test for independence for ordinal data (chap 2.5 in the text, and Handout 6).

Friday, Oct 19 Fisher's exact test (chap 2.6 in the text and Handout 7), three-way contingency tables (chap 2.7 in the text and Handout 7).

Homework 3: Due on Monday, October 22

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. Consider the following contingency table, in which two species of mice were tested for a specific parasite.

	Infected	Not Infected
Species 1	20	8
Species 2	12	17

(a) If π_1 and π_2 are the proportions of the infected in species 1 and species 2 respectively, then find an approximate confidence interval for $\pi_1 - \pi_2$. Interpret your result.
(b) Estimate the odds of infection for the two species (two numbers). Also find the odds ratio comparing species 1 to species 2.
(c) Obtain an approximate 95% confidence interval for the odds ratio. Interpret your result.

2. Consider the following contingency table (based on the past record of 100,000 individuals), in which the true diagnosis (by an invasive test) and the predicted diagnosis (by a mouth swab) are shown.

	Tested Positive	Tested Negative
Condition True	475	25
Condition False	4975	94525

(a) Estimate the overall risk (probability) of having the condition.
(b) Estimate the sensitivity and specificity.
(c) Estimate the false negative (having the condition given that the test is negative), and false positive (do not have the condition given that the test is positive).
(d) Would you say that this test does well? Explain your answer.

3. The following table is obtained from a General Social Survey in 2002.

	Party Identification		
	Democrat	Independent	Republican
White	871	444	873
Black	302	80	43

(a) Carry out a test for the null hypothesis of independence between party identification and race. Use $\alpha = 0.05$. Find the p-value of your test and state your conclusion.
(b) Obtain the standardized residuals and summarize your findings.
(c) Partition the chi-squared into two components, and explain your results.

4. In a large country, smoking behavior of a random sample of 709 lung cancer patients were recorded. Independently, a random sample of 709 lung cancer free patients was taken and their smoking behavior were recorded. The data is from 1950. The counts are given below.

Have smoked	Cancer	Cancer-free
Yes	688	650
No	21	59
Total	709	709

(a) Identify the response variable and the explanatory variable. Describe the sampling scheme. Is it joint multinomial or independent multinomials? If it a joint multinomial, are the rows independent multinomials or are the columns independent multinomials. Explain
(b) Can you obtain the proportion who have lung cancer in the country? Explain.
(c) Estimate the odds ratios of smoking among the cancer, and among the cancer free patients. Estimate the odds ratio θ which compares the odds of cancer to cancer-free patients. Summarize your findings.
(d) Test the hypothesis $H_0 : \theta \leq 1$ and against $H_1 : \theta > 1$ at level $\alpha = 0.05$. Find the p-value of your test and state your conclusion.

5. In an $I \times J$ contingence table, let $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$ be the estimated expected counts under H_0, the hypothesis of independence.
(a) Show that the row and column totals of $\{\hat{\mu}_{ij}\}$ are the same as those of the observed counts $\{n_{ij}\}$.
(b) When the rows are independent multinomials, then show that $E(\hat{\mu}_{ij}) = E(n_{ij})$ under H_0.
(c) Show that $E(n_{+j}/n) = \pi_{+j}$ under H_0, irrespective of whether $\{n_{ij}\}$ are joint multinomial or the rows of $\{n_{ij}\}$ are independent multinomials.
(d) If the rows of $\{n_{ij}\}$ are independent multinomials, and if $\{\pi_{i+}\}$ were known, then show that $\tilde{\pi}_j = \sum_{i=1}^{I} n_{ij}\pi_{i+}/n_{i+}$ is unbiased for π_{+j} (even when H_0 is false). .