

1. For the following statements, answer true(T) or false(F). (2 pts. each)
 - a. _____ In 2×2 tables, statistical independence is equivalent to a population odds ratio value of $\theta = 1.0$.
 - b. _____ A British study reported in the *New York Times*: (Dec. 3, 1998) stated that of smokers who get lung cancer, “women were 1.7 times more vulnerable than men to get small-cell lung cancer.” The number 1.7 is a sample odds ratio.
 - c. _____ Using data from the Harvard Physician’s Health Study, we find a 95% confidence interval for the relative risk relating having a heart attack to drug (placebo, aspirin) to be (1.4, 2.3). If we had formed the table with aspirin in the first row (instead of placebo), then the 95% confidence interval would have been $(1/2.3, 1/1.4) = (.4, .7)$.
 - d. _____ Pearson’s chi-squared test of independence treats both the rows and the columns of the contingency table as nominal scale; thus, if either or both variables are ordinal, the test ignores that information.
 - e. _____ For testing independence with random samples, Pearson’s X^2 statistic and the likelihood-ratio G^2 statistic both have chi-squared distributions for any sample size, as long as the sample was randomly selected.
 - f. _____ Fisher’s exact test is a test of the null hypothesis of independence for 2×2 contingency tables that fixes the row and column totals and uses a hypergeometric distribution for the count in the first cell. For a one-sided alternative of a positive association (i.e., odds ratio > 1), the P-value is the sum of the probabilities of all those tables that have count in the first cell at least as large as observed, for the given marginal totals.
 - g. _____ The difference of proportions, relative risk, and odds ratio are valid measures for summarizing 2×2 tables for either prospective or retrospective (e.g., case-control) studies.
 - h. _____ In a 5×2 contingency table that compares 5 groups on a binary response variable, the G^2 chi-squared statistic with $df = 4$ for testing independence can be exactly partitioned into 4 separate independent chi-squared statistics that each have $df = 1$ by comparing row 1 to row 5, row 2 to row 5, row 3 to row 5, and row 4 to row 5.
2. (5 pts.) Silicon wafers for computer chips manufactured by a high-tech company have an average number of defects of 1 per wafer. If the distribution of the number of defects follows the Poisson distribution, find the probability that a wafer has *at least* 1 defect.

3. (9 pts.) Each of 100 multiple-choice questions on an exam has five possible answers but one correct response. For each question, a student randomly selects one response as the answer.

a. Specify the probability distribution of the student's number of correct answers on the exam, identifying the parameter(s) for that distribution.

b. Would it be surprising if the student made at least 50 correct responses? Explain your reasoning.

4. (10 pts) Consider the following data from a women's health study:

		Myocardial Infarction	
		Yes	no
Oral Contraceptives	Used	23	34
	Never Used	35	132

a. Construct a 95% confidence interval for the population odds ratio.

b. Based on (a), does it seem plausible that the variables are independent? Explain.

5. (12 pts.) For adults who sailed on the Titanic on its fateful voyage, the odds ratio between gender (female, male) and survival (yes, no) was 11.4.
- What is wrong with the interpretation, “The probability of survival for females was 11.4 times that for males.”
 - When would the quoted interpretation be approximately correct? Why?
 - The odds of survival for females equaled 2.9. For each gender, find the proportion who survived.
6. (13 pts.) In a General Social Survey, gender was cross-classified with party identification. Table 1 attached shows some results.
- Explain what the second set of numbers are in the contingency table. Show how to obtain 261.42.
 - Explain how to interpret the Prob value listed for the Chi-Square statistic.
 - Explain carefully how to interpret the values listed under StReschi. In which

Table 1: Table for Problem 6

Frequency							
Expected dem indep repub							
-----+-----+-----+-----+							
female 279 73 225							
261.42 70.653 244.93							
-----+-----+-----+-----+							
male 165 47 191							
182.58 49.347 171.07							
-----+-----+-----+-----+							

Statistic	DF	Value	Prob

Chi-Square	2	7.0095	0.0301
Likelihood Ratio Chi-Square	2	7.0026	0.0302

Observ	Resraw	Reschi	StReschi	Observ	Resraw	Reschi	StReschi
1	17.584	1.088	2.293	4	-17.584	-1.301	-2.293
2	2.347	0.279	0.465	5	-2.347	-0.334	-0.464
3	-19.931	-1.274	-2.618	6	19.931	1.524	2.618

cases were there significantly more people than one would expect if party identification were independent of gender?

Formulas

Poisson $P(y) = e^{-\mu} \mu^y / y!$, $y = 0, 1, 2, 3, \dots$

Binomial $P(y) = \frac{N!}{y!(N-y)!} \pi^y (1-\pi)^{N-y}$, $y = 0, 1, 2, \dots, N$, $\mu = N\pi$, $\sigma = \sqrt{N\pi(1-\pi)}$

Hypergeometric $P(n_{11}) = \frac{\binom{n_{1+}}{n_{11}} \binom{n_{2+}}{n_{+1}-n_{11}}}{\binom{n}{n_{+1}}}$, $\binom{a}{b} = a! / [b!(a-b)!]$

odds = $\pi / (1 - \pi)$, $\pi = \text{odds} / (1 + \text{odds})$, relative risk = π_1 / π_2

$$\theta = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)}, \quad \hat{\theta} = \frac{n_{11} n_{22}}{n_{12} n_{21}}$$

$$\text{ASE}(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

$$X^2 = \sum \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}, \quad \hat{\mu}_{ij} = (n_{i+} n_{+j}) / n, \quad df = (I - 1)(J - 1)$$

$$G^2 = 2 \sum n_{ij} \log \left(\frac{n_{ij}}{\hat{\mu}_{ij}} \right), \quad df = (I - 1)(J - 1)$$

$$r_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - p_{i+})(1 - p_{+j})}}$$

$$\text{CMH} = \frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k \text{Var}(n_{11k})}, \quad df = 1$$

$$\pi = \alpha + \beta x, \quad \log[\pi / (1 - \pi)] = \alpha + \beta x$$