Due by 4pm on Friday, April 4. Please leave your homework on the table before class begins on Friday or leave it in the dropbox outside my office. Do not forget to attach the honor code. There are five problems worth 20 points each.

(1) Last year, Instacart used their shopping data to determine the most popular pie in each state. Suppose that we take a small sample of residents from CT and MA and ask them whether they prefer apple or blueberry pie. Results are tallied below:

	Blueberry	Apple	Total
MA	16	4	20
CT	2	10	12
Total	18	14	32

(a) Suppose that we want to use a uniform prior to estimate the proportion of people who prefer blueberry pie in each state. What is the posterior distribution for MA? What is the posterior distribution for CT?
(b) Construct a 95% CI for the odds ratio comparing blueberry preferences in MA to CT. Use R.
(c) Construct a 90% credible interval for the difference in the proportion of people who prefer blueberry pie in MA and CT. Use R.
(d) What is the approximate probability that the proportion of people in MA who prefer blueberry pie is larger than that of in CT? Use R.

(2) Does whether or not a student brings a car to campus depend on how far away they live? A survey of HC students collected the following information and whether or not students had a car on campus and their home state.

	Yes	No	Total
MA	19	29	48
Non-MA	17	33	50
Total	36	62	98

(a) What is the odds ratio of having a car on campus if you live in MA (compared to those who don't live in MA)? Interpret your odds ratio in the context of the problem.
(b) Construct a 95% CI for the odds ratio of having a car on campus depending on whether you live in-state or out of state.
(c) At the 10% level of significance, test for independence between a student's home state and whether or not they have a car on campus.

(3) A survey was conducted asking people about their opinion about the company's proposed new health plan and their perceived level of job stress. Results are tabulated below:

Job Stress	Favorable	Unfavorable	Total
Low	250	750	1000
High	400	1600	2000
Total	650	2350	3000

(a) Test whether or not a person's opinion about the company's proposed new health plan is associated with / dependent upon their perceived level of job stress. Calculate both the χ^2 and G^2 test statistics.
(b) Which cell(s) contribute the most to the χ^2 statistic? Interpret your findings in the context of the problem.

(4) A random sample of 216 patients having a certain skin disease were classified by age (recorded as categories)

	Age Category				
Severity	1	2	3	4	Total
Moderate	15	32	18	5	70
Mildly Severe	8	29	23	18	78
Severe	1	20	25	22	68
Total	24	81	66	45	216

(a) Is the severity of the skin disease independent of age? Calculate both the χ^2 and G^2 test statistics.
(b) Which cell contributes the most to the test statistic? Interpret your result in terms of the severity of the skin disease and age.
(c) Interpret the standardized residuals for cells. Use R.

(5) Extracorporeal membrane oxygenation (ECMO) is a potentially life-saving procedure for treating newborn babies suffering from severe respiratory failure. An experiment was conducted in which 29 babies were treated with ECMO and 10 babies treated with conventional medical therapy (CMT).

	Die	Live	Total
ECMO	1	28	29
CMT	4	6	10
Total	5	34	39

(a) What are the largest and smallest possible values for n_{11} in this problem?
(b) Use Fisher's Exact test to find the p-value associated with a test to claim that ECMO is more effective than CMT. Use R to compute the sum of hypergeometric probabilities.
(c) What is the p-value for the test if we instead used the mid-p-value approach?
(d) Use Fisher's Exact test to find the p-value associated with a test to claim that ECMO and CMT are not equally effective. Use R to compute the sum of hypergeometric probabilities.