- The final exam is scheduled to be on **Friday, September 15th 4:00 pm – 7:00 pm** at HSSB 1173 and a possible second overflow room (see the Gauchospace for the updates & room assignment).
- No cell-phones, no laptops, no tablets, no electronic readers, no devices with the cellular or internet access, etc.
- The exam is closed-book, but you're allowed to bring with you:
  - a two-sided page of *your own* handwritten notes (Please do not copy notes from your classmates. You should go through the material on your own and write down what you think would be helpful on the exam.);
  - the Table of Distributions from Gauchospace website or from the textbook (no other tables are allowed);
  - a calculator;
  - your ID.
- Cheating will not be tolerated. Anyone caught cheating will receive an automatic F on the exam. In addition the incident will be reported, and dealt with according to University's Academic Dishonesty regulations. Please refrain from talking to your peers, exchanging papers, writing utensils or other objects, or walking around the room. All of these activities can be considered cheating.

Below are some additional problems for you to study. (The problems were provided by Professor Drew Carter.) These questions cover the material from the second half of the course. You should also review the midterm, midterm review and the Bayesian inference questions from the last assignment for additional relevant material.

1. The grades in an upper division class are tabled to compare the performance of juniors and seniors:

	A	B	C	F
Juniors	16	32	23	2
Seniors	8	24	16	3

   We want to test whether or not the distribution of grades was the same for the two classes of students.

   (a) Is the chi-squared approximation appropriate for this table? If necessary, redraw the table in such a way that it corrects the problem.

   (b) Calculate the appropriate test statistic for the χ^2 test of independence.

   (c) What is the critical value for this test with size $\alpha = 0.01$?

   (d) What do you conclude?

2. The following table counts the number of outcomes in 100 trials for each of four events:

Event	A	B	C	D
Outcomes	34	30	12	24

Perform goodness-of-fit test of size $\alpha = 0.05$ to test the hypothesis that $P(A) = P(B)$ and $P(C) = P(D)$.

3. The chair of the statistic department is interested in the majors of students in the 120 courses. The registrar tabulated the number of engineers, economist, and statisticians in each semester of 120.

	Engineering	Economics	Statistics
120A	112	63	35
120B	80	53	27
120C	40	23	12

Test whether or not the distribution of majors is the same for the three classes.(Use size $\alpha = 0.05$) What interpretation would you give to the results?

4. The National Traffic and Safety Administration reported the following data on fatal accidents in Los Angeles County over 3 years. In each accident two factors were recorded: whether alcohol was involved or not and whether speeding was involved or not. Here is a cross tabulation:

	2003	2004	2005
Neither Speeding nor Alcohol	241	251	232
Speeding Only	244	243	235
Alcohol Only	259	206	223
Both Speeding and Alcohol	68	53	60

(a) Suppose that instead of using the null hypothesis that the marginal distribution of the two factors is independent of the year (as we did in a homework problem), we wanted to test whether the conditional probability of speeding did not change from year to year. What would be the expected number of accidents in 2004 where speeding was involved conditional on the fact that alcohol was involved under the null hypothesis that the conditional distribution (given alcohol) is the same for each year?

(b) The χ^2 test statistic for this test is $X^2 = 0.25$. Is this significant?

(c) Interpret your conclusion.

5. A study of juror summonses recorded the race/ethnicity of 1271 people sent jury summonses, and then tracked how many of them served on juries, how many had their service deferred, and how many never showed up.

	No Show	Deferred	Served
White	438	250	39
Black	103	120	12
Hispanic	187	115	7

We want to test the null hypothesis that the outcome of their jury service is independent of race and ethnicity.

(a) Calculate the expected number of people in each of the cells in the third row of the table (i.e. how many Hispanics do we expect to be no shows, to be deferred, and to serve on a jury, respectively) assuming the null hypothesis is true.

(b) The χ^2 test statistic for this table is $X^2 = 26.9$. What do you conclude? (Use $\alpha = 0.01$)

6. In the same survey as in question 6, the sample contained 625 men and 646 women.

Men

	No Show	Deferred	Served
White	225	133	20
Black	47	52	7
Hispanic	87	52	2

Women

	No Show	Deferred	Served
White	213	117	19
Black	56	68	5
Hispanic	100	63	5

We want to test the hypothesis that their jury service is marginally independent of gender and race.

(a) Calculate the expected number Hispanic women that served on a jury and the expected number of Hispanic men that served on a jury.

(b) Is the chi-squred approximation appropriate for this table? If not, then how can it be fixed?

(c) If the test statistic is $X^2 = 28.75$, then what do we conclude at an $\alpha = 0.01$ level?

7. A model for incomes in a heterogeneous population takes n independent observations $X_1, \ldots, X_n$ and transforms them to generate $Y_i = \log X_i$.

Assume that Y_i have independent normal distributions with unknown mean θ and variance $\sigma^2 = 1$. We will impose a prior distribution on this mean parameter that is normal with mean 10 and variance 5.

(a) For 38 observations with $\bar{y} = 11.129$, what is our Bayes estimator of θ?

(b) Give a 95% Bayesian Credible Interval for θ.

(c) To generate an estimate on the original scale we need e^θ. Calculate the Bayes estimator for e^θ.

8. The daily returns on a portfolio managed by Fiduciary Inc. were recorded over 275 days. The average return was 0.0038 with a sample standard deviation of 0.00534. Then for the next 12 days, a new strategy was tested, and it produced an average return of 0.0067 with a sample standard deviation of 0.00832.

(a) Perform a two-sample t test to see whether there is a significant improvement in the strategy over the old. What would you conclude at a $\alpha = 0.05$ level?

(b) What assumptions are you making in part a? How can you check those assumptions? Be specific.

9. An environmental engineer is studying the effects of a sand bar on the flow rate through 15 coastal lagoons. The engineer measures the flow rate in each lagoon before and after the sand bar is built, and the increase in the rate is recorded in gallons per minute (gpm). The engineer has a prior that the average difference in the flow rate is normally distributed with an expectation of 0 and a standard deviation of 75 gpm.

(a) Find the Bayesian estimator of the average difference in the flow rate if we observe that the 15 lagoons saw an average increase of 5.86 gpm. (We can assume that the measurements come form a normal distribution with standard deviation 25 gpm.)

(b) Give a 95% credible interval for the average flow rate.

(c) In a sentence or two, explain to the engineer what this interval represents.

(d) Calculate the posterior probability that the average flow rate is greater than 0.

10. A survey of 356 registered voters in Carpinteria asked them their opinions on the local ballot measures (Measure K and Measure J) in todays election. The respondents were divided into three groups for each question: supporters, opposition, and those that said they dont know or wont be voting. Here is cross-tabulation of the responses:

		Measure J		
		Support	Oppose	Don't Know
	Support	86	92	15
Measure K	Oppose	40	25	17
	Don't Know	36	23	22

(a) Suppose that we are only interested in the outcome of Measure J. Would we conclude that there is significant difference in the proportions of supporters versus opposition to Measure J? Calculate the appropriate test statistic and interpret your result (use $\alpha = 0.05$).

(b) For testing whether the respondents answers to the two questions are independent, the expected number of respondents under the null hypothesis are tabled below.

		Measure J		
		Support	Oppose	Don't Know
	Support	87.83	??	??
Measure K	Oppose	37.31	32.25	12.44
	Don't Know	36.86	31.85	??

Calculate the 3 missing expectations.

(c) For the test from part (b), Calculate a test statistic. What would we conclude at a $\alpha = 0.05$ level?

11. A study on the causes of ulcers took a random sample of patients with peptic ulcers or gastric ulcers as well as a control group of healthy patients. The blood type (O, A, or B) and gender of each patient was recorded.

Male Patients

	Peptic	Gastric	Control
O	472	190	1470
A	285	208	1356
B	62	43	250

Female Patients

	Peptic	Gastric	Control
O	511	193	1422
A	394	208	1269
B	72	41	320

The researchers want to test if the ulcer diagnosis group is independent of gender and blood type. We want to use a null hypothesis of marginal independence.

(a) Calculate the three corresponding expected values for Female patients with Type O blood.

(b) Confirm that the test statistic is $X^2 = 63.40$. What do we conclude?

12. A union supervisor claims that applicants for jobs are selected without regard to race. The hiring records of the local – one that contains all male members – gave the following sequence of White (W) and Black (B) hirings:

WWWWBWWWBBWBB

Do these data suggest a nonrandom racial selection in the hiring of the union's members?