POS 3713
Spring 2001
Review Problems for Final Exam

General Information: The final exam is on Friday, April 27th from 10:00am-12:00pm. It will consist of 5 multiple choice questions and several short answer/problem-solving questions. The exam will cover the following chapters from Healey (10-18), and Chapter 17 in Manheim & Rich. You will be responsible for material only since the last exam. This includes Analysis of Variance (ANOVA), Chi square test for independence, Phi, Lambda, Gamma, Tau-b, Pearson's r (correlation), and Regression (calculation of the slope and y-intercept for bivariate regression and interpretation of results for multivariate regression including t-tests, R^2 and F). You should be able to conduct a hypothesis test for each of these statistics. We will go over the following review problems in class on Wednesday, April 18th and Friday, April 20th. You can find solutions to these problems on Dr. Mitchell's webpage (http://www.geocities.com/CollegePark/Plaza/3094/). Good luck!

1. The Tallahassee City government is interested in attracting large manufacturing plants to locate within the city. They hire Dr. Bailey to study the factors that influence how many manufacturing plants are located within a city. After doing a preliminary review of past research, he decides that the number of businesses located in a city depends on the business tax rate and the amount of money the city spends on infrastructure (roads, schools, government agencies to aid businesses, etc.) In particular, Dr. Bailey argues that the higher the business tax rate, the *lower* the number of businesses that will locate in a city. Furthermore, he believes that the greater the amount of money a city spends on infrastructure, the *higher* the number of businesses that will locate in that city.

   He collects the following data on 31 cities around the country: (a) the number of manufacturing plants locating within the city over the last five years, (b) the average tax rate on industry over the same time period (a percentage with a possible range between 0 and 100%), and (c) the total dollars spent by the city on infrastructure improvement over that time period (in thousands of dollars).

   Regression Results
Variable	Estimated Coefficient (b)	Standard Error (s.e.)
Intercept	5.621624	2.610154
Tax Rate	-1.494506	0.625466
Infrastructure	0.001716	0.000666

   $N = 31$ $R^2 = 0.648$ $F = 5.45$ (p value = .01)

   a. Write down the appropriate null and alternative hypotheses for Dr. Bailey's model. Do the results support his theoretical hypotheses?
   b. What is the substantive impact of the tax rate and money spent on infrastructure on the number of businesses that locate within a city? In Dr. Bailey's model, what

impact does a one percent increase in the tax rate have on the number of businesses that locate in a city? What impact does a one unit increase in money spent on infrastructure (1 unit = $1000) have on the number of businesses that locate in a city?

c. What does the reported R^2 tell you about the overall fit of Dr. Bailey's model? What does the reported F-statistic tell you about the overall fit of the model? Does his model do a good job in explaining the number of businesses that locate in a city?

d. The city of Tallahassee wants to use these results to make a prediction of the number of plants that will locate in our city in the next five years based on this regression model. The tax rate in Tallahassee is 5% (tax rate = 5), while the money spent by the city on infrastructure in the past 5 years is $5 million (infrastructure = 5000). How many businesses do you expect to locate in Tallahassee based on this information?

2. Dr. Mitchell and Dr. Prins are interested in studying the issues that democratic nations fight about. They believe that well established democracies (those democracies whose institutions are strong and have been around for quite a long time) behave differently than new democracies (with new institutions that can potentially be challenged). In particular, they assert that new democracies are more likely to fight over territorial issues than older, more established democracies. They collect the following data on militarized disputes to test their hypothesis in the Post World War II period (1946-1992).

Democracy Level	Non-territorial Disputes	Territorial Disputes	Total
New Democracies	43	31	74
Established Democracies	21	2	23
Total	64	33	97

$$\chi^2 = 8.61$$

a. Conduct a χ^2 test for independence. State the null and alternative hypotheses. Assume a 95% level of confidence, i.e., $\alpha = .05$. Do your results support Mitchell and Prins theoretical hypothesis that newer democracies are more likely to fight over territorial issues?

b. Calculate ϕ (phi) and interpret your results. Is this consistent with what you found in (a)?

3. A recent National Election Study of American citizens includes a question about how interested in politics the respondents perceive themselves to be (ranging from low to medium to high). You would like to determine if the respondents' education level has influenced their level of interest in politics. In particular, you would like to know if education in general improves a person's level of education about and interest in politics (which is usually called civic education). The survey results are presented below.

Count

		Highest level of education			
		Grade School	High School	College	Total
Interest in Politics	Low	78	69	37	184
	Medium	25	43	34	102
	High	15	40	76	131
Total		118	152	147	417

a. Calculate the value of gamma for this table. How do you interpret this number? Does it support the contention that higher levels of education in general promote greater civic education? In other words, are people with more education more likely to be interested in politics?

b. Conduct a statistical hypothesis test for the value of Gamma reported in (a). Suppose that the calculated t statistic equals 3.772 with a p-value less than .0001. State the null and alternative hypotheses. What can you conclude about the relationship between education and interest in politics based on this result?

4. In an attempt to prepare for a 5-K race, Bob decides that he should come up with a training strategy. Besides smoking a few cigarettes and eating little chocolate donuts, Bob decides that for a week he will record how long it takes him to run practice 5-K's. During his practice, Bob becomes thoroughly convinced that the more chocolate donuts he eats, the faster is he able to run. Below is a list of the number of chocolate donuts he ate before 7 practices and the corresponding recorded time. Use this information to answer the following questions below.

# of Chocolate Donuts Consumed (X)	Recorded Running Time (Y)
3	35.10
5	34.50
7	33.02
10	29.36
13	28.11
17	27.27
20	26.30

a) Calculate the regression line.
b) How do you interpret the coefficient for X (i.e., b), the number of chocolate donuts Bob eats?
c) If Bob eats 9 donuts, how fast can he expect to run?
d) Calculate the correlation, or Pearson's r, between the # of donuts consumed and the recorded running time. Interpret your results. In other words, is Bob's conclusion that the more donuts he eats, the faster he runs supported?

5. A random sample of 30 countries has been rated as predominantly urban, suburban, or rural. You are interested in determining if the infant mortality rate (number of infant deaths per 1000 live births) in these countries varies significantly by the level of urbanization. The ANOVA results are presented below.

Level of Urbanization	Mean Mortality	Infant Mortality	Sum of squares	df
Rural	11.77	Between Groups (SSB)	41.45	2 (k-1)
Urban	9.47	Within Groups (SSW)	164.098	27 (N-k)
Suburban	9.12	Total (SST)	205.548	29 (N-1)

a. State the null and alternative hypotheses.

b. Compare the means. Do they seem different just by looking at the mean infant mortality rate (you do not need to do a test here, just tell me if the mean levels look different)?

c. Test your hypothesis for a difference in means with analysis of variance (ANOVA), i.e., conduct an F test assuming $\alpha = .05$. Does the infant mortality rate (number of infant deaths per 1000 live births) in these countries vary significantly by the level of urbanization?

		Party Identification		
		Democrat	Republican	Independent
	Jewish	19	11	0
Religion	Catholic	35	15	10
	Protestant	10	50	10

6. Calculate λ (Lambda) for the data in the table above (assume party identification is the dependent variable). What does Lambda tell us about the relationship between religion and party identification, in other words, how do you interpret your results? If you know that the calculated t value for this statistic is 3.231, with a p-value of .001, can you conclude that religion has a significant impact on party identification (i.e., is it significantly different from zero)?