Problems: HYPOTHESIS TESTING

1. Research on the impact of "test anxiety" on test performance is inconclusive. Some theorists argue that test anxiety increases students' concentration and thus improves their performance on tests. Other theorists argue that test anxiety is a distraction for students and thus impedes test performance. Your research seeks to evaluate which (if either) of these theories is true.

The average final exam grade is 80% for all students who took STAT 401 during the five years prior to this past summer. During the past summer semester you exposed the 12 students in a STAT 401 course to a new technique for administering tests. To reduce their test anxiety below its level during previous STAT 401 final exams, you had the students stand, stretch their arms, and take a deep breath before beginning to work on the exam. Scores on the exam are as follows:

77% 85% 73% 70% 74% 65% 83% 78% 82% 72% 68% 73%

a. Give an estimate of central tendency and an estimate of dispersion based on these final exam scores. (Show how your estimates were derived and give units for each estimate.)

b. What is the final exam grade at the 35th percentile?

c. What are your null and alternative hypotheses?

d. Do your data provide statistically significant evidence (at the .05 level of significance) for either theory? If so, which one? (Justify your answer.)

e. How large a sample would have been needed to have detected a one percent (1%) increase in final exam scores as statistically significant at the .05 significance level? (Hint: You will want to use an estimate of the population variance from the above data in answering this question.)

2. When the government of Carlos Salinas Gortari took office in 1987, Mexico had but one billionaire to its name. After seven years of the Salinas government's neoliberal economic policies, it had (in July of 1994) 24 billionaires—more than all countries of the world except the U.S., Germany, and Japan. Yet Mexico's crushing rural and urban poverty remained, leaving the "35 families" (i.e., the Mexican ruling class) richer, with 65-80% of Mexico's wealth, and leaving other Mexicans to scramble for the remaining 20-35%.

Neoliberal economic theory suggests that a country rich in natural resources and cheap labor (such as Mexico) should be a prime target for foreign investment. Once foreigners invest in such a country, employment should rise, wages and standards of living should improve, and a consequently growing middle class should strengthen the country's economy. But this did not happen in Mexico, and you think that you know why. You believe that Mexican employers kept their work forces earning subsistence wages by not allowing labor unions to form and by firing anyone who complained of not earning
enough. In other words, your theory is that no Mexican middle class emerged because employers repressed employees' attempts to reduce their poverty. Wealth was siphoned off to the rich and the chasm between rich and poor grew rather than shrank.

To evaluate your theory you obtain data from Nuevo Laredo, a Mexican border town with 15 unionized U.S. factories. You wish to evaluate whether these unionized workers' incomes are greater than the median annual income among all Mexicans (currently 6,231 pesos, which is less than $1,000 at the current exchange rate). During interviews with the chief union representative from each of the 15 factories you obtain the following data on the median income (in thousands of pesos per year) among the workers in each factory:

6.5 5.5 7.9 7.6 6.7 7.4 8.6 6.8 8.3 7.4 7.7 5.0 7.1 6.3

Be sure that you read these numbers correctly: The first number listed (i.e., 6.5) indicates that 6,500 pesos is the median annual income among the workers of the first factory.

a. Give a 95% confidence interval for the average workers' median income (i.e., the mean among the median incomes) among the factories. Show your work.

b. Give null and alternative hypotheses that would allow you to test your theory (as described in the opening paragraphs of this problem).

c. Using the .05 significance level find the critical value(s) that delineates the acceptance region for the hypothesis test specified in part b.

d. Given your theoretical orientation, what is the P-value associated with the average median income that you found among the 15 unionized factories in Nuevo Laredo?

e. State in words the meaning of the P-value found in part d. (Hint: The P-value is a number. Construct a sentence that uses this number in a way that demonstrates your knowledge of what a P-value is.)

3. According to the 1990 census, the proportion of poor residents (with incomes below the poverty line) of Southern Detroit was .75. You want to test whether poverty has increased since then (i.e., whether there is a smaller proportion of nonpoor residents there now), but you have almost no money and can only randomly sample five people.

a. State your null and alternative hypotheses.

b. Determine whether you could detect an increase in the proportion of poor residents at the .05 level with your sample of five. Explain your answer. (Hint: In doing this problem, you will want to calculate the probability that all five residents in your sample are poor.)

c. How large would your sample have to be to detect an estimated increase of more than .10 in the proportion of poor residents at the .05 level of significance?
4. Distinguish between the $P$-value and the significance level associated with a hypothesis test.

5. Four years ago the governor signed a bill into law that increased the taxes on pig farmers by an additional five dollars per pig. Your agency has been asked to determine whether this tax increase has undermined (i.e., decreased) farmers' confidence in the profitability of pork production in your state. Your belief is that farmers will stop raising pigs, if they believe that the tax makes pork production unprofitable. Prior to the signing of the bill, a census of all farmers in the state showed that farmers owned 6.7 pigs on average. The census also showed the standard deviation in pigs-per-farmer to be 8.4 pigs.

   a. If you were to obtain data on each farmer's current number of pigs (based on a random sample of farmers in your state), what null and alternative hypotheses would you use to evaluate whether in your state the tax undermined farmers' confidence in the profitability of pork production?

   b. At the .05 significance level, how large a sample would be required for you to obtain a precision of 2 pigs?

   c. You obtain data on a random sample of 78 farmers in your state. You find that the farmers in your sample own 4.0 pigs on average, and that the standard deviation in pigs-per-farmer is 8.0 pigs. Based on these data, find the appropriate rejection rule, and test the hypotheses given in part a. Explain in plain English (i.e., NOT in statistical jargon) what conclusion can be drawn from this test. (Hints: Use the .05 significance level, and remember to use your current data to estimate the standard error of the sampling distribution.)

   d. What is the $P$-value of your findings?

6. Of the declarations of war that nondemocratic countries made during the last fifty years, 60% of them were declarations of war on countries that had already declared war on them (that is to say, 60% were defensive declarations of war). You theorize that democratic countries are more likely than this to have declared war defensively. Your sample is of nine situations in which democratic countries declared war. You obtain data on a variable that takes the following two values:

   1 = war was declared defensively by the democratic country

   0 = war was not declared defensively by the democratic country

   a. What is the unit of analysis in this research?

   b. State your null and alternative hypotheses.

   c. What would be your rejection rule at the .05 significance level?

   d. Assume for the balance of this problem that in eight of nine situations in your sample "war was declared by a democratic country on a country that
had already declared war on it" (i.e., for each of the eight situations assume a value of 1 on your variable).

1) Give a point estimate for the mean on your variable. (Show your work.)

2) What is the attained significance (or P-value) of this estimate? (Show your work.)

3) Does the evidence support your theory at the .05 level of significance? (Explain your answer.)

7. The Rural Mental Health Center (RMHC) does research on family distress in Iowa farm families. You have obtained from the Center a list of the names of all 200 families-in-distress from among their random sample of 1000 Iowan farm families. Your second data source is from the records of Iowa MasterCard, Inc. Searching through these records you find that 10 of the families-in-distress (from the 200 in the RMHC list) have MasterCard credit cards. Moreover, only 1 (or 10%) of these 10 regularly pays its credit card balance every month. This is considerably less than the 42% of all Iowan farm families with MasterCard credit cards, who regularly pay their credit card balances every month.

Your theory is that family distress is likely to find expression in other aspects of families' lives, such as in money mismanagement. For this reason you hypothesize that families-in-distress will be more likely to mismanage their money than families in general. You adopt the RMHC's definition of family distress and designate money-mismanagement-families to be families that do not regularly pay their credit card balances each month.

a. State the null and alternative hypotheses that you would use to test your theory.

b. In testing this hypothesis, you will use data drawn from a specific population. What is this population?

c. What is the sampling distribution of your test statistic under the null hypothesis? Be as exact as possible in your answer.

d. At the .05 significance level, perform the hypothesis test appropriate to the hypotheses given in part a. State your conclusion in clear English (i.e., not in statistical jargon).

e. What is the P-value associated with your finding that only one family in your sample regularly pays its credit card balance every month?

f. Are the findings consistent with your theory? Explain your answer.

8. During the last municipal election the Los Angeles Police Commissioner asserted to an audience at the steps of city hall, "Youth gang members do not listen to reason; force is the only language they understand." His law-and-order stance was popular with the voters and he soon became the Los Angeles Mayor. As his first official action the newly elected mayor doubled the Los
Angeles police force and tripled the number of police patrols within neighborhoods overrun by youth gangs.

The new L.A. Police Commissioner wishes to deploy his police in the way that most effectively will reduce gang warfare (i.e., violence between rival youth gangs). His theory is that more gang warfare occurs "at the borders between" rather than "in the interiors of" territories claimed by distinct gangs. His thinking is that most gang warfare is due to differences in the territories that rival gangs believe are "theirs," and that it is when rival gang members find each other in these disputed territories that they will seek to reestablish territorial rights by acting violently toward each other. To test his theory he hires you to investigate whether fewer gang-related homicides occur in neighborhoods in which police patrol only the borders of gang territories, and not all parts of these territories.

You begin your research by using informers to determine the borders of territories controlled by the various L.A. youth gangs. As you learn more about the gangs you discover that the gangs have business specialties in drugs, prostitution, and/or gambling. Because other research suggests that drug dealers are more likely to be violent than those who run gambling or prostitution businesses, you decide to include "business specialty" as a variable in your initial analysis. Your data are on all 68 gang-related homicides in L.A. during the last six months.

Table 1: Gang-related Homicides Occurring Within Border- Versus Interior-Patrolled Territories Controlled by Youth Gangs Dealing in Drugs, Prostitution, or Gambling

<table>
<thead>
<tr>
<th>Business Specialty</th>
<th>Border-patrolled</th>
<th>Interior-patrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>drugs</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>prostitution</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>gambling</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>more than one</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

Be sure that you read this table correctly: The "6" in the table's upper left-hand cell means that in Los Angeles six gang-related homicides occurred during the last six months in border-patrolled territories that are controlled by exclusively drug-dealing gangs.

a. Based on the data in Table 1 would you conclude that for gang-related homicides in Los Angeles neighborhoods, the border- versus interior-patrolling method used in the territory where it occurred is independent of the business specialty of the gang that controls the territory? Use the .05 significance level and show your work.

b. Give an estimate of the conditional probability that a gang-related homicide occurs in a border-patrolled territory of Los Angeles given that the territory is controlled by exclusive drug dealers (i.e., by drug dealers who deal neither in prostitution nor gambling). Show your work!

c. Estimate the joint probability that a gang-related homicide occurs in a border-patrolled territory of Los Angeles that is controlled by exclusive drug dealers and prostitution.
drug dealers. Show your work! (Hint: Do NOT assume that patrolling technique and business specialty are independent.)

d. What is the marginal probability that a gang-related homicide occurred in a border-patrolled territory of Los Angeles? Show your work! (Hint: Note that in this part as well as in parts e, f, and g you must ignore all data in Table 1 related to gangs' business specialties.)

e. Set up null and alternative hypotheses for testing whether among gang-controlled territories of Los Angeles, fewer gang-related homicides occur in border-patrolled territories than in interior-patrolled territories. (Hint: Keep in mind that the proportion of border-patrolled territories plus the proportion of interior-patrolled territories must equal one.)

f. Using the .05 significance level state the rejection rule needed to test the hypotheses set up in part e. (Hint: Answering this part will require that you obtain a variance estimate using the data in Table 1.)

g. Test the hypotheses listed in part e (i.e., apply the rejection rule obtained in part f). Note evidence that supports your conclusion. Also be sure to explain your conclusion in clear English, not in statistical jargon.