Case Problem 1: National Health Care Association(Descriptive Statistics)

The National Health Care Association is concerned about the shortage of nurses the health care profession is projecting for the future. To learn the current degree of job satisfaction among nurses, the association has sponsored a study of hospital nurses throughout the country. As part of this study, a sample of 50 nurses was asked to indicate their degree of satisfaction in their work, their pay and their opportunities for promotion. Each of the three aspects of satisfaction was measured on a scale from 0 to 100, with larger values indicating higher degrees of satisfaction. The data collected also showed the type of hospital employing the nurses. The types of hospitals were private (P), Veterans Administration (VA) and University (U). The complete data set is on the file named "Health".

Managerial Report:

Use methods of descriptive statistics to summarize the data. Present the summaries that will be beneficial in communicating the results to others. Discuss your findings. Specifically, comment on the following questions.

1. On the basis of the entire data set and the three job satisfaction variables, what aspect of the job is most satisfying for the nurses? What appears to be the least satisfying? In what area(s), if any, do you feel improvements should be made? Discuss
2. On the basis of descriptive measures of variability, what measure of job satisfaction appears to generate the greatest difference of opinion among the nurses? Explain
3. What can be learned about the types of hospitals? Does any particular type of hospital seem to have better levels of job satisfaction than the other types? Do your results suggest any recommendations for learning about and/or improving job satisfaction? Discuss
4. What additional descriptive statistics and insights can you use to learn about and possibly improve job satisfaction?

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 2: Hamilton County Judges(Probability)

Hamilton County judges try thousands of cases per year. In an overwhelming majority of cases disposed, the verdict stands are rendered. However, some cases are appealed, and of those appealed, some of the cases are reversed. Kristen DelGuzzi of the *The Cincinnati Enquirer* conducted a study of cases handled by Hamilton County judges over the years 1994 through 1997 (*The Cincinnati Enquirer,* January 11, 1998). Shown in Data File "Judge" are the results for 182,908 cases handled (disposed) by 38 judges in Common Pleas Court, Domestic Relations Court and Municipal Court. Two of the judges (Dinkelacker and Hogan) did not serve in the same court for the entire three-year period.

The purpose of the newspaper's study was to evaluate the performance of the judges. Appeals are often the result of mistakes made by judges and the newspaper wanted to know which judges were doing a good job and which were making too many mistakes. You have been called in to assist in the data analysis. Use your knowledge of probability and conditional probability to help with the ranking of the judges. You also may be able to analyze the likelihood of cases handled by different courts being appealed and reversed.

Managerial Report:

Prepare a report with your rankings of the judges. Also, include an analysis of the likelihood of appeal and case reversal in the three courts. At a minimum your report should include the following:

1. The probability of cases being appealed and reversed in the three different courts

2. The probability of a case being appealed for each judge

3. The probability of a case being reversed for each judge

4. The probability of reversal given an appeal for each judge

5. Rank the judges within each court. State the criteria you used and provide a rationale for your choice.

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 3: Bock Investment Services(Estimation)

Lisa Rae Bock started Bock Investment Services (BIS) in 1994 with the goal of making BIS the leading money market advisory service in South Carolina. To provide better service for her present clients and to attract new clients, she has developed a weekly newsletter. Lisa has been considering adding a new feature to the newsletter that will report the results of a weekly telephone survey of fund managers. To investigate the feasibility of offering this service, and to determine what type of information to include in the newsletter, Lisa selected a sample random sample of 45 money market funds. The Data File "Bock", reports fund assets and yields for the past seven and 30 days. Before calling the money market fund managers to obtain additional data, Lisa decided to do some preliminary analysis of the data already collected.

Managerial Report

1. Use appropriate descriptive statistics to summarize the data on assets and yields for the money market funds.
2. Develop a 95% confidence interval estimate of the mean assets, mean 7-day yield, and mean 30-day yield for the population of money market funds. Provide a managerial interpretation of each interval estimate
3. Discuss the implication of your findings in terms of how Lisa could use this type of information in preparing her weekly newsletter.
4. What other information would you recommend that Lisa gather to provide the most useful information to her clients?

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 4: Quality Associates Inc (Hypothesis Testing Univariate Population)

Quality Associates Inc., is a consulting firm that advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. In one particular application, a client gave Quality Associates a sample of 800 observations taken during a time in which that client's process was operating satisfactorily. The sample standard deviation for these data was 0.21; hence the population standard deviation was assumed to be 0.21 Quality Associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. By analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. When the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. The design specification indicated the mean for the process should be 12. The hypothesis test suggested by Quality Associates follows:

$$H_o: \mu = 12$$
$$H_1: \mu \neq 12$$

Corrective action will be taken any time H_o is rejected.

Samples were collected at hourly intervals during the first day of operation of the new statistical process control procedure. These data are available in the data set "Quality".

Managerial Report:

1. Conduct the hypothesis test for each sample at the 0.01 level of significance and determine what action, if any, should be taken. Provide the test statistic and p value for each teach test.
2. Compare the standard deviation for each of the four samples. Does the assumption of 0.21 for the population standard deviation appear reasonable?
3. Compute limits for the sample mean $\bar{X}$ around μ such that, as long as a new sample mean$\bar{X}$ is within those limits, the process will be considered to be operating satisfactorily. If $\bar{X}$ exceeds the upper limit or if $\bar{X}$ is below the lower limit, corrective action will be taken. These limits are referred to as upper and lower control limits for quality control purposes.
4. Discuss the implications of changing the level of significance to a larger value. What mistake or error could increase if that were done?

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 5: Par Inc(Hypothesis Testing Bivariate Populations)

Par Inc., is a major manufacturer of golf equipment. Management believes that Par's market share could be increased with the introduction of a cut-resistant, longer-lasting golf ball. Therefore, the research group at Par has been investigating a new golf ball coating designed to resist cuts and provide a more durable ball. The tests with the coating have been promising.

One of the researchers voiced concern about the effect of the new coating on driving distances. Par would like the new cut-resistant ball to offer driving distances comparable to those of the current-model golf ball. To compare the driving distances for the two balls, 40 balls of both the new and current models were subjected to distance tests. The testing was performed with a mechanical hitting machine so that any difference between the mean distances for the two models could be attributed to a difference in the design. The results of the tests, with distances measured to the nearest yard, are contained in the data set "Golf".

Managerial Report:

1. Formulate and present the rationale for a hypothesis test that par could use to compare the driving distances of the current and new golf balls
2. Analyze the data to provide the hypothesis testing conclusion. What is the p-value for your test? What is your recommendation for Par Inc.?
3. Provide descriptive statistical summaries of the data for each model
4. What is the 95% confidence interval for the population mean of each model, and what is the 95% confidence interval for the difference between the means of the two population?
5. Do you see a need for larger sample sizes and more testing with the golf balls? Discuss

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 6: A Bipartisan Agenda for Change (Tests of Goodness of Fit and Independence)

In a study conducted by Zogby International for the *Democrat and Chronicle*, more than 700 New Yorkers were polled to determine whether the New York state government works. Respondents surveyed were asked questions involving pay cuts for the state legislators, restrictions on lobbyists, terms limits for legislators and whether state citizens should be able to put matters directly on the state ballot for a vote (*Democrat and Chronicle*, December 7, 1997). The results regarding several proposed reforms had broad support, crossing all demographic and political lines.

Suppose that a follow-up survey of 100 individuals who live in the western region of New York was conducted. The party affiliation (Democrat, Independent, Republican) of each individual surveyed was recorded, as well as their responses to the following three questions.

1. Should legislative pay be cut for every day the state budget is late
   Yes ------- No -------
2. Should there be more restrictions on lobbyists?
   Yes ------- No -------
3. Should there be term limits requiring that legislators serve a fixed number of years?
   Yes ------- No -------

The responses were coded using 1 for a *yes* response and 2 for a *no* response. The complete data set is available on the data disk in the data set name "NYReform".

Managerial Report:

1. Use descriptive statistics to summarize the data from this study. What are your preliminary conclusions about the independence of the response (yes or no) and party affiliation for each of the three questions in the survey?
2. With regard to question 1, test for the independence of the response (yes or no) and party affiliation Use $\alpha = 0.05$
3. With regard to question 2, test for the independence of the response (yes or no) and party affiliation Use $\alpha = 0.05$
4. With regard to question 3, test for the independence of the response (yes or no) and party affiliation Use $\alpha = 0.05$
5. Does it appear that there is broad support for change across all political lines? Explain

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 7: Wentworth Medical Centre(ANOVA)

As part of a long-term study of individuals 65 years of age or older, sociologists and physicians at the Wentworth Medical Centre in update New York investigated the relationship between geographic location and depressions. A sample of 60 individuals, all in reasonably good health, was selected; 20 individuals were residents of Florida, 20 were residents of New York and 20 were residents of North Carolina. Each of the individuals sampled was given a standardized test to measure depression. The data collected follow; higher test scores indicate higher levels of depression. These data are available on the data disk in the file "Medical1".

A second part of the study considered the relationship between geographic location and depression for individuals 65 years of age or older who had a chronic health condition such as arthritis, hypertension, and/or heart ailment. A sample of 60 individuals with such conditions was identified. Again, 20 were residents of Florida, 20 were residents of New York, and were residents of North Carolina. The levels of depression recorded for this study follow. These data are available on the data disk in the file "Medical2".

Managerial Report:

1. Use descriptive statistics to summarize the data from the two studies. What are your preliminary observations about the depression scores?
2. Use analysis of variance on both data sets. State the hypotheses being test in each case. What are your conclusions?
3. Use inferences about individual treatment means where appropriate. What are your conclusions?
4. Discuss extensions of this study or other analyses that you feel might be helped.

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 8: Compensation for ID Professionals(ANOVA with Interaction)

For the last 10 years, *Industrial Distribution* has been tracking compensation of individual distribution (ID) professionals. Results for the 358 respondents in the 1997 Annual Salary Survey showed that 27% of the respondents work for companies with sales over $40 million, with the typical ID professional working for a $12 million firm. Those who work for small to mid-sized companies (between $6 million and $20 million) report higher earnings than those in larger firms. The lowest paid employees work for firms with sales of less than $1 million. The typical outside salesperson made $50,000 in 1996 and the typical inside salesperson earned $30,000 (*Industrial Distribution*, November 1997). Suppose that a local chapter of ID professionals in the greater San Francisco area conducted a survey of its membership to study the relationship, if any, between the years of experience and salary for individuals employed in outside and inside sales positions. On the survey, respondents, were asked to specify one of three levels of years of experience: low (1-10 years); medium (11-20 years) and high (21 or more years) The complete data set consisting of 120 observations, is available on the data file "IDSalary"

Managerial Report:

1. Use descriptive statistics to summarize the data
2. Develop a 95% confidence interval estimate of the mean annual salary for all salespersons, regardless of years of experience and type of position
3. Develop a 95% confidence interval estimate of the mean salary for outside salespersons. Compare your results with the national value reported by *Industrial Distribution*
4. Develop a 95% confidence interval estimate of the mean salary for inside salesperson. Compare your results with the national value reported by *Industrial Distribution*
5. Ignoring the years of experience, develop a 95% confidence interval estimate of the mean difference between the annual salary for outside salespersons and the mean annual salary for inside salespersons. What is your conclusion?
6. Use analysis of variance to test for any significant differences due to position. Use a 0.05 level of significance and for now, ignore the effect of years of experience
7. Use analysis of variance to test for any significant differences due to years of experience. Use a 0.05 level of significance and for now ignore the effect of position.
8. At the 0.05 level of significance test for any significant differences due to position, years of experience and interaction. Use inferences about individual treatment means where appropriate.

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 9: Spending & Student Achievement(Simple Linear Regression)

Is the educational achievement level of students related to how much the state in which they reside spends on education? In many communities this important question is being asked by taxpayers who are being asked by their school districts to increase the amount of tax revenue spent on education. In this case, you will be asked to analyze data on spending and achievement scores in order to determine whether there is any relationship between spending and student achievement in the public schools.

The federal government's National Assessment of Education Progress (NAEP) program is frequently used to measure the educational achievement of students. Table 14.13 shows the total current spending per pupil per year and the composite NAEP test score for 35 states that participated in NAEP program. These data are available on the data file named "NAEP".

Managerial Report:

Develop numerical and graphical summaries of the data
Use regression analysis to investigate the relationship between the amount spent per pupil and the composite score on the NAEP test. Discuss your findings.
Do you think that the estimated regression equation developed for these data could be used to estimate the composite scores for the states that did not participate in the NAEP program?
Suppose that you only considered states that spend at least $4,000 per pupil but not more than $6,000 per pupil For these states, does the relationship between the two variables appear to be any different than for the complete data set. Discuss the results of your findings and whether you think deleting states with spending less than $4,000 per year and more than $6,000 per pupil is appropriate.
Develop estimates of the composite scores for the states that did not participate in the NAEP program.
Based upon your analyses, do you think that the educational achievement level of students is related to how much the state spends on education?

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)

Case Problem 10: Consumer Research Inc(Multiple Regression)

Consumer Research Inc. is an independent agency that conducts research on consumer attitudes and behaviors for a variety of firms. In one study, a client asked for an investigation of consumer characteristics that can be used to predict the amount charged

by credit card users. Data were collected on annual income, household size and annual credit card charges for a sample of 50 consumers. the file named "Consumer" contains the relevant information.

Managerial Report:

1. Use methods of descriptive statistics to summarize the data. Comment on the findings
2. Develop estimated regression equations, first using annual income as the independent variable and then using household size as the independent variable. Which variable is the better predictor of annual credit card charges? Discuss your findings
3. Develop an estimated regression equation with annual income and household size as the independent variables. Discuss your findings
4. What is the predicted annual credit card charge for a three-person household with an annual income of $40,000?
5. Discuss the need for other independent variables that could be added to the model. What additional variables might be helpful?

(This case problem is due to Anderson, Sweeney, and Williams for Classroom Discussion)