BASIC DOCTORAL WRITTEN EXAMINATION IN BIOSTATISTICS
PART II

May 8 – May 15, 2001
(PICKUP at 12:00 NOON)

INSTRUCTIONS

a) This is an open-book "take home" examination.

b) Answer any four (but only four) of the six questions which follow.
NOTE: There are six questions to choose from *this year only* because of the special circumstances of the recent change in the curriculum. The policy that only five questions will be offered has not been changed.

c) Put the answers to different questions on separate sets of papers.

d) Since your papers may be xeroxed for back-up purposes, type or write them with a paper and pen/pencil combination that will xerox clearly. Do not, for example, use a hard pencil on yellow paper.

e) Most questions should be answered in the equivalent of less than five typewritten pages (300 words per page), and under no circumstances will more than the first 10 typewritten pages or the equivalent be read by the grader.

f) Put your code letter, not your name, on each page.

g) Return the examination with a signed statement of the Honor Pledge:

"In recognition of and in the spirit of the honor code, I certify that I have neither given nor received aid on this examination and that I will report all Honor Code violations observed by me."

(Signed)_____
NAME

NOTE:
All the computer files to which this examination refers are available on Departmental computers in the directory O:\BIOSLIB\BASICS\~~2000\~~2001\. The examination itself is 2001PT2.~~DOC~~PDF.

Question 1:

Enclosed are relevant portions of the design description and selected analysis findings from a statewide survey of migrant farm workers in the state of Colorado. Topics covered in this material include: (i) motivation and general purpose, (ii), overview of access to care, (iii) information on the Migrant Health System, (iv) statement of the research problem, (v), description of research plan (including descriptions of the sample design, questionnaire development, and the data collection plan), and (vi) a summary of the analysis findings on utilization of health care services.

Please prepare a concise critique of aspects of the design and analysis raised below. Support your discussion with results from the statistical literature, wherever it is appropriate to do so. If in the course of preparing your responses you feel there are important design details that are missing from this document, specifically state any assumptions that you need to complete your discussion.

Points

7 a. **Sample Design** -- First summarize the design; i.e., for each stage of sampling indicate the following: the sampling unit, how (if at all) stratification is used, the type of selection method used to randomly choose sampling units in the stage, and the total number of sampling units to be chosen in that stage. Then critique the choice of sampling units, the number of sampling stages, and the use of stratification and/or clustering. You should point out any strengths and weaknesses in the general approach. If you believe there are reasonable alternatives to what has been done, briefly describe and discuss the relative merits of the existing and alternative strategy.

4 b. **Sampling Frame** -- Discuss the quality of the sampling frames that would have been used for this sample. Comment specifically on the following: undercoverage, the presence of population non-members on the frame, and frame multiplicity.

4 c. **Sample Weights** -- Explicitly defining all of your notation, provide the formula that would have been used to compute the selection probability and unadjusted sample weight for members of the sample. What, if any, adjustments to these weight would have been needed? Briefly explain why.

6 d. **Precision of Estimates** -- The formula for estimating the variance of estimates from this sample was purposely excluded from the background information that is provided. Using the estimated percent of the population reporting their health as "good" (p. 21) to illustrate, see if you can figure out what formula would have been used to estimate the variance of this estimated percentage (95% confidence intervals in parentheses). Also explicitly indicate how you would have set up SUDAAN to properly account for this sample design.

4 e. **Other Nonsampling Aspects of the Research Design** -- Briefly sketch the nonsampling error implications (i.e., due to measurement problems and nonresponse) in this design. Based on your assessments of these and the sampling aspects of the study design, do you think the study met its research objectives?

ACCESS TO HEALTH CARE: A SURVEY OF COLORADO'S MIGRNT FARMWORKERS

(Narrative Published in International Migration Review, Vol. 21, Fall, 1987, Center For Migration Studies)

Carla N. Littlefield, Ph. D.
Charles L. Stout, M.P.H.

Colorado Migrant Health Program

SEPTEMBER, 1987

INTRODUCTION

Migrant farmworkers are among the most deprived of the nation's medically indigent. Pervasive poverty, underemployment, isolation, and alienation make their access to health services a critical issue. In the current climate of reallocations and program reductions, it is important to assess the government's fulfillment of its ethical and moral obligation to provide adequate care to this disadvantaged minority.

The President's Commission for the study of Ethical Problems in Medicine and Biomedical and Behavior Research (1983:4-5) concluded that:

> Society has an ethical obligation to ensure equitable access to health care for all...Equitable access to health care requires that all citizens be able to secure an adequate level of care without excessive burdens...The Federal government has the ultimate responsibility for seeing that health care is available to all when the market, private charity, and government efforts at the state and local levels are insufficient in achieving equity...Efforts to contain rising health care costs are important but should not focus on limiting the attainment of equitable access for the least well served portion of the public.

The purpose of this report is to utilize data collected in a sample survey of Colorado's adult migrant farmworker population to determine their health needs, health services utilization, and overall access to care. Health needs include selected indices of medical, dental, nutrition and reproductive health. The conclusions and recommendations of the report address pertinent issues in the funding and delivery of health care services to the migrant farmworker population.

ACCESS TO HEALTH CARE

Current efforts to limit federal health care expenditures threaten to minimize the gains made in the 1960s and 1970s to increase access to health care by traditionally disadvantaged groups (Aday et al., 1984; Andersen et al., 1986). Access is defined as "those dimensions which describe the potential and actual entry of a given population group to the health care system" (Aday et al., 1984:13). Two broad influences on access are

characteristics of both the health care system and the population at risk. According to the framework proposed by Aday & Andersen (1975) and Aday et al. (1984), the health care system's influence is exerted through its structure, resources, and policies. For the population at risk, characteristics related to health needs, wants, resources, utilization, and satisfaction with available health services are associated with access to services.

Reviewing the literature and utilizing data from several national surveys, Andersen et al. (1986) applied this framework to access of health care by Hispanics, the predominant migrant farmworker group in Colorado. Significant characteristics of the population include education, occupation, income, language, health care beliefs and attitudes, regular source of health care, insurance coverage, and health needs. The relationship between low education and reduced access to care may be attributed to traditional health beliefs and practices, a distrust of modern medicine, or lack of information about available services. Low occupational status is often associated with both low income and limited or no insurance. Language differences often pose a formidable barrier to both accessing and utilizing health services in facilities without bilingual staff.

Perceiving their health status to be generally less favorable than that reported by white non-Hispanics (NRA, 1980), Hispanics are also less likely to have a usual source of health care (Carter, 1985). Service utilization rates of Hispanics are also lower than for whites, notably in the areas of hospital visits, dental visits (NCHS, 1984), prenatal care, and family planning (NCHS, 1981). Finally, the cultural beliefs of many Hispanics include health practices which may place more reliance on home remedies, over-the-counter medicines, herbs, and teas as a response to both acute and chronic disease conditions. Confidence in folk healers as a substitute or supplement to the modern health care system is a dominant theme for a significant minority of migrants in making personal health decisions.

Focusing on the health system, Andersen et al. (1986:242) reviewed the Hispanic health literature for criticisms of the health care system and noted, "language barriers, middle-class values and attitudes, and cultural disparities between client and providers." Concomitantly, dissatisfaction with service was higher among Hispanics than other ethnic groups, especially in the areas of cost, waiting time, interaction with the physician, and time spent with the physician.

More difficult to measure are the differences in health and illness beliefs, understandings, and expectations of ethnic minorities and those of health professionals in the modern health care system, differences which may pose a barrier for care for the minority client (Sakala, 1985). While in traditional cultures illness is viewed as a disruption in the harmony or balance between an individual and the social, spiritual, or physical environment, for modern medicine the focus is on the clinical assessment of signs and symptoms. These differences in conceptualization of health and illness demand increased communication between provider and client but the structure, values and resources of the health care system often function to thwart this goal.

THE MIGRANT HEALTH SYSTEM

Federal Congress established funding for migrant health and community health services in 1975 and finalized the regulations in 1977. A migratory agricultural worker is defined as an individual whose principal employment is in agriculture on a seasonal basis, who has been so employed within the last 24 months, and who establishes a temporary place of abode for such employment. A migrant health center or program is defined as an entity which either through its staff and supporting resources or through contracts or cooperative arrangements with other public or private entities provides for migratory agricultural workers, seasonal agricultural workers, and the members

of the families of such workers. Funded services include primary care, preventive medical services, preventive dental services, emergency health care and supplemental services such as hospitalization, dental, vision, pharmacy, health education, and outreach.

By 1987, 130 migrant health centers provided health care to over 230,000 migrant farmworkers annually with a federal budget of $44.5 million. Approximately 85 percent of this amount is allocated for primary care operations, approximately five percent is administrative and approximately 10 percent is discretionary. It is significant to note that 10 percent descretionary spending is a major amount during a period of level funding , inflation, and increased demand for services.

Unfortunately, there is little documentation that those in greatest need of services are receiving them. The Office of Migrant Health in the Department of Health and Human Services (the principal national funding resource for migrant health programs) estimates that only 12-17 percent of the eligible population are recipients of care through its national system. Access to health care is an issue which migrant health administrators and advisers question on a continuing basis (Migrant Health Task Force, 1986, 1987). The need to demonstrate outcome indicators of access, other than dollars spent, is more crucial now than ever before, particularly with planning assumptions that forecast level funding for the next two to three years.

Colorado Colorado is one of over 30 "upstream" states with federally funded programs, upstream indicating that the state is a destination for migrants seeking temporary residence to engage in agricultural labor. The Colorado Migrant Health Program (CMHP) plays a lead role in the provision of health services to a targeted population of approximately 43,000 migrant and seasonal farmworkers and dependent family members. Operating as a section of the Colorado Department of Health, CMHP annually serves approximately 7,000

medical patients and 3,000 dental patients, including nearly 2,000 school-age children enrolled in migrant education summer schools. These services are provided in a variety of settings: outreach clinics, community-based health care organizations, private health care providers with CMHP agreements, migrant education schools, migrant Daycare/Headstart centers and one county health department. An additional combined total of 9,500 medical patients and 3,600 dental patients are served annually by three other migrant health grantees (two in North Central Colorado and one in the San Luis Valley).

THE PROBLEM

Although Colorado has a well-developed statewide system of coordinated health care services for migrant and seasonal farmworkers, effective health care planning has been hampered by the absence of current information on the migrant population's access to health care. Specifically lacking has been basic data on population demographics (age, sex, education, language, family size, work history, permanent residence), health status, health services utilization, and user satisfaction with services received.

The overall goal of the study was to survey Colorado's adult male and female migrant farmworker population, ages 18-50 years, for the purpose of evaluating access to health care and developing recommendations for improving delivery of migrant health care. Related objectives were to 1) develop a research tool (questionnaire) in Spanish and English which could be used in replicating the survey elsewhere in the United States and 2) develop a research methodology which could be replicated elsewhere to generate a scientifically designed random sample of the migrant farmworker population. It is hoped that through similar studies in both upstream and downstream states, an accurate profile of this country's migrant farmworker population will be developed. It is further hoped that once a profile which establishes health needs and access to health care is developed, that it will then be used as a basis for decision-making regarding funding allocations and the deployment of limited resources nationally.

METHODS

Research Design The research design was developed in conjunction with onsite consultation from the National Center for Health Statistics. This consultation incorporated consideration into the sample design of the unusual constraints posed by a migrant population. These constraints included: 1) uncertain estimates for migrant population totals in each area; 2) uncertain predictions as to which housing units would be occupied in a given area; 3) high mobility of the population present in the state during the interviewing period; 4) vagaries in the weather which precipitate unanticipated shifts in employment prompting the migrants to move to other areas or to other states.

For these sampling conditions, it was decided that the most effective approach would be a descriptive study based on a stratified, proportionate, probability sample of the migrant farmworker population residing in Colorado between July 1, 1986 and September 30, 1986. The two strata identified were 1) families and 2) solo males. Solo males were defined as males migrating alone or with other males, and who reside with other males; their marital status is not necessarily "single". The sampling was proportionate in order to account for variations in the numbers of solo males and families estimated to reside in the four areas of the state selected for inclusion in the study. Those areas were: the north central, southeast, south central, and Western Slope, all identified by the CMHP and the Colorado Department of Agriculture as being the most labor-intensive areas of agriculture in the state.

Sample Design The eligible population for the survey was persons 18-50 years of age who were active migrant farmworkers at the time of the survey. The selection of persons for this survey was determined through a scientifically designed sample. A multistage stratified proportionate probability design was employed to arrive at the sample. The State of Colorado was stratified into

four regions, identified above, where there was a known high concentration of migrant farmworkers based on information from previous years. Each of the four strata was divided into two substrata. One substratum consisted of the latest available listing of all family housing units (FHU's) within a particular stratum. The other substratum consisted of a listing of all single male beds within all dormitories within a particular stratum.

An attempt was made at arriving at a self-weighting sample, that is, an effort was made in the sampling scheme in selecting persons with approximately the same probability. This was accomplished by applying the overall sampling fraction (desired total number of sample persons divided by the approximate number of eligible persons in the population) to the estimated number of eligible persons within each substratum. The number of adults in the FHU substrata was approximated by multiplying the estimated average number of FHU's in that substratum by a factor of 3 (assumed to be the average number of adults per family in all regions, except in the Western slope where two adults per family was assumed). Within the solo male substrata, the overall sampling fraction was simply applied to the estimated number of solo males in each substratum.

Once the sample number of FHU's and sample number of solo males were determined within a stratum, the actual selection was carried out by taking a systematic random sample of FHU's and dormitory bed numbers within each substratum. Dormitory beds were selected in a single stage process. However, the selection of sample persons within the FHU strata was carried out as a two-stage process. That is, the FHU was selected in the first stage and then eligible males and females were listed separately within the selected FHU. The second stage of selection consisted of randomly selecting a male and a female from the sample FHU. If only one male or female resided in the sample FHU, each was selected with certainty (probability = 1).

Field Procedures Related to Random Selection of Respondents No listing of the target population was available from which to select a sample prior to initiation of the survey. Instead, lists of the addresses or locations of residences usually occupied by migrants were constructed for each identified area of the state, one list for family housing units (FHU's) and one list for solo male dormitories. Listers then investigated the solo male dormitories to determine the number of beds available; these bed numbers then comprised the lists used to designate the solo male respondents. Although some residences for solo males were vacant at the time of the listings, it was assumed that they would be occupied at some time during the study period and their bed numbers were added to the solo male lists.

From the two separate lists (FHU's and bed numbers in the solo male dormitories), the sample for the study was randomly selected. The assigned lists with the designated FHU's and solo male dormitory bed numbers were given to each interviewer along with a detailed map of the specific geographical area. In each designated FHU interviewers listed and numbered all adult males and all adult females eligible for inclusion in the study. One adult male and one adult female was then chosen randomly (numbered chips selected blindly from a bag) for the interview. In each dormitory the solo males occupying the designated numbered beds were the designated solo males for the interviews. No substitutions were allowed. Following the interview, each respondent was given five dollars. After completion of the dental screening, the respondent was given an additional five dollars.

In each area there were fewer occupied residences than anticipated, for reasons primarily related to the weather. The severe spring frost on the Western Slope damaged the fruit crops to the extent that only half the work force was hired for the summer and fall harvests. Severe hail damage in the

vegetable and sugar beet fields in the north central area closed two family camps early because of lack of work. One camp for solo males in the south central area failed to open, without any clear reason known to anyone.

Instrument Development Two instruments were developed for the Colorado Migrant Health Survey: 1) a household screening form to determine eligible family members in sample housing units, and 2) a survey questionnaire in Spanish and English. The questions related to Health Services, Hypertension, Pesticide Exposure, and Medicine/Vitamin Usage were derived from the Hispanic Health and Nutrition Examination Survey (NCHS 1985). The family planning section was based almost entirely on the Centers for Disease Control questionnaire used in their U.S.- Mexico border survey (1979). Several questions related to hunger were adapted from a tool developed by the Food Research and Action Center (FRAC 1986) and the Utah Nutrition Monitoring Project (1985). The Selected Conditions list was derived from the Wisconsin migrant farmworker survey (Slesinger, 1979).

Interviewer Training Fourteen bilingual interviewers, five males and nine females, were hired on a part-time basis to work throughout the state through the end of September. Male interviewers were used exclusively for interviews with males, and female interviewers were used predominantly for interviews with females. All interviewers were required to complete an intensive two-day training program, after which they were given their assignments. The first three completed surveys of each interviewer were carefully reviewed to ensure that instructions were followed. In addition to weekly telephone contact, the field supervisor made at least two trips to each area to supervise the interviewers and monitor their progress in the completion of their assignments.

Dental Screening. Dental screening was carried out at a pre-arranged time after the respondent completed the interview, either in a local dental clinic or at the migrant site. Eight dental examiners, dental hygienists and dentists, were individually trained to conduct two standardized dental indices: a dental caries status and treatment index (DMFT-modified) and a periodontal health index (CPITN). Dental examiners used calibrated examination instruments and approved portable examination equipment. Dental screening times were scheduled for evening hours to accommodate field work schedules; transporation to and from the screening locations was provided by the survey interviewers when needed. The interviewers, present at every screening, provided a familiar link between the participant and the dental examiner and intrepreted when necessary.

Data Collection Results The original goal of the study was to complete 600 interviews. Because of the weather, several camps remained closed throughout the summer and at least two closed earlier than usual. The subsequent listings resulted in 513 designated respondents of which 331 were interviewed, a response rate of 64.5 percent. Of the 182 nonparticipants, there were 52 refusals; 93 moved after one eligible respondent in the FHU (Family Housing Unit) had been interviewed and included in the study; 37 were ineligible because they did not meet the study criteria (i.e. one family adult met the study criteria but the other family adult was not within the age range or was not a farmworker).

Dental screening was completed on 172 respondents. Of the 159 interviews without dental screening, there were 46 refusals; 46 persons moved prior to dental screening. One area was without dental examiners resulting in 67 non-dental screenings.

Data Analysis and Statistical Computations Data analysis was performed on a Compaq microcomputer using the Statistical Package for the Social Sciences (SPSS/PC). The statistics presented are based on a sample of the target population rather than on the entire population. The difference between an estimate based on a sample and the true population is called the sampling error. The expected magnitude of the sampling error is measured by a statistic called the standard error. The standard error is used to compute the confidence interval, the estimate plus or minus two standard errors of the estimate within which the true population value lies with 95 percent confidence. The computation of confidence intervals and standard errors was adjusted to account for the difference between the targeted sample and actual sample in each stratum. The following equations taken from Kish (1965) provide a correction and describe the calculations. (attached).

THE POPULATION

Interviews from a total of 329 adult farmworkers, between the ages of 18 and 50 years, were analyzed. In terms of ethnicity, Hispanics predominated (94.2%) followed by American Indian (2.4%) Anglo (1.2%) or other (2.1%). The sample was stratified by solo males and families resulting in the following totals: 129 family males, 126 family females, and 74 solo males. There were differences among the three groups in terms of age, marital status, place of permanent residence, language skills, and whether this was their first visit to Colorado. The term "sex status" is used when differentiating family males, family females and solo males. Confidence intervals are indicated in parentheses, substituted with an asterisk when a small N resulted in a confidence interval greater than the estimate value.

UTILIZATION OF HEALTH SERVICES

As a moving population, predominantly monolingual in Spanish, migrant farmworkers face many potential obstacles related to access to care. This section looks at health status, usual place of care, and problems encountered while seeking care. The incidence of hospitalizations, days in bed, and farmworker injuries are documented.

Health Condition Over half (50.5%) identified their health as fair or poor (Table 19). The other half (49.5%) reported being in good, very good, or excellent health. Findings were similar to those reported in the Wisconsin migrant worker survey (Slesinger, 1979), but reflect their inferior perceived health status relative to the U.S. population as a whole (NCHS, 1984)

Table 19
Health Status Comparison, Colorado Migrant and U.S. Adult Population
Numbers and Percentages

	Colorado		NHS**
	Numbers	Percentages	Percentages
Excellent	27	8.2 *	44.0
Very Good	24	7.3 *	27.9
Good	112	34.0 (5.7)	21.5
Fair	140	42.6 (4.9)	6.1
Poor	26	7.9 *	NA
Total	329	100.0	100.0

**National Health Survey, (NCHS, 1984). Poor and fair were combined in the NHS report.

Usual Place of Care One fourth of the population (25.3%, N=83) did not have a usual place of health care at their permanent residence, slightly higher than the 22.4 percent reported by Mexican Americans and considerably higher than the 14.1 percent reported by Non-Hispanic whites in the preliminary Hispanic HANES report (Carter et al. 1985). For the total sample (N=329), Table 20 displays the places identified. Women (83.3%) were more likely than

men (69.5%) to identify a usual place of care and were more likely (44.4%) to use the migrant clinic than either family males (23.3%) or solo males (4.1%)

Table 20
Usual Place of Health Care at Permanent Residence, Numbers and Percentages

	Number	Percentage
Migrant Clinic	89	27.2 (5.7)
Community or Rural Health Center	44	13.4 (8.4)
Other Clinic	25	7.6 *
Hospital Outpatient Clinic	39	11.9 (8.8)
Doctor's Office/Clinic	39	11.9 (8.6)
Hospital Emergency Room	7	2.1 *
Other Place	2	0.6 *
No Usual Place of Health Care	83	25.3 *
Total	328	100.0

Of the total sample (N=329), one third (34.2% N=105) utilized a federally-sponsored health facility (migrant health clinic, community health center or health clinic) as a usual place of health care at their permanent residence. However, one fourth ((25.4% N=78) had no usual source of care and one fifth (22.1% N=68) had a permanent residence outside the U.S. (Mexico or Guatemala) where subsidized care may or may not have been available.

Table 21
Utilization of Federally Sponsored Health Care at Permanent Residence, Numbers and Percentages**

	Number	Percentage
No Usual Source of Care	78	25.4
Usual Source of Care - Mx., Guat.	68	22.2
Homebase Texas - Source of Care Non-Federally sponsored	47	15.3
Homebase Other U.S. State - Source Of Care Non-Federally sponsored	9	2.9
Homebase Texas - Source of Care Federally sponsored	80	26.1
Homebase Other State - Source of care Federally sponsored	25	8.1
	307	100.0

**Excludes all respondents whose homebase is Colorado.

More than one half (51.3% N=168) of the migrants interviewed identified Texas as their place of permanent residence. Nearly half (47.6% N=80) of those from Texas reported using federally sponsored resources (migrant health clinic, community health or other health clinics) as a usual source of care (Table 22). However, nearly one fourth of the Texas-based migrants (24.4% N=41) indicated they had no usual source of care.

Table 22
Usual Place of Health Care - Texas Homebase

	Number	Percentage
Migrant Health Clinic	57	33.9
Community Health Center	11	6.6
Other Clinic	12	7.1
Hospital Outpatient Clinic	27	16.1
Dr. Office/Private	15	8.9
Hospital - E.R.	5	3.0
No Usual Place of Health Care	41	24.4
	168	100.0

Use of Health Services in Colorado Over half (52.6% N=173) of the population had used health services in Colorado at some time, and women were more likely than males to report this (chi square <.05). Forty percent (N=133) of the total population identified their usual place of care in Colorado as the migrant health clinic, community health center or other clinic (Table 23). This is congruent with the percentage of migrants using federally-sponsored health services either in Texas or other states. However, it is not known how many of those interviewed subsequently accessed these services in Colorado after the interview process. It is important to note again that in 1986, 34 percent of the adult migrants in Colorado were working and living in the state for the first time.

Table 23
Usual Place of Health Care in Colorado,
Numbers and Percentages

	Number	Percentage
Migrant Health Clinic	128	38.9
Community Health Center	3	0.9
Other Clinic	2	0.6
Hospital Outpatient Clinic	8	2.4
Doctor's Office/Clinic	14	4.3
No Usual Place of Health Care	18	5.5
Never Used Colorado Services	156	47.4
	329	100.0

One third (33.8% N=52) of the migrants who had a usual place of health care in Colorado had heard about the place from the migrant health program outreach worker. The other referrals were distributed fairly evenly among friends (18.2%), relatives (14.9%), crew leaders (11.7%), or other (12.3%). Fourteen (9.1%) had found the place themselves.

Last Medical Visit Of the total respondents, twelve (3.6%) had never had a medical visit. Forty-five percent (N=148) had not received health care in the previous year, a larger proportion than reported for Mexican Americans (31.0 %) or non-Hispanic whites (28.3%) in the preliminary Hispanic HANES report (Carter, 1985). Women (74.6%) were more likely than males (42.9%) to have had a medical visit in the past year. Of the entire sample, one fourth (25.6%) had had their last visit one to five years ago. For 55 persons (17.4%), this visit was over five years ago or they couldn't remember when it occurred.

Those who reported having had a medical visit (N=317) identified where the visit had occurred. Forty-three percent (N=136) had seen a provider in Colorado; 32.6 percent (N=103) had had this visit at their permanent residence. Thirty-eight (12%) had seen a provider in some other state. The

remainder (12.4%) had taken place outside the United States. One third of those who could recall their last medical visit related it to illness (36.6%), followed by injury (12.6%), checkup (9.8%), pregnancy care (9.1%) or other (31.9%).

Satisfaction With Care Of those who had a previously medical visit, ninety-four percent (N=294) said they got the care they wanted. The majority (63.1%) stated they were very satisfied with the care received. A small percentage (8.7% N=27) were "not at all satisfied". The other 28 percent (N=88) were somewhat satisfied. Males, especially solo males, were more likely (chi square $p < .05$) to report dissatisfaction with the care received.

Although voicing satisfaction with the care received, over one fourth of those who had a previous medical visit had experienced a problem. The problems identified are summarized in Table 24 in rank order of frequency. Women were more likely to report that the doctor didn't diagnose or treat the condition. Both women and solo males were more likely (chi square $<.05$) to report mistreatment by the doctor or staff than were family males. Whether miscommunication was a factor is not known.

Table 24
Source of Dissatisfaction with Last Medical Visit,
Frequencies and Percentages*

	Frequency	Percentage
Condition didn't improve after treatment	85	27.3 (6.7)
Doctor didn't diagnose or treat condition	77	24.8 (6.9)
Not enough time with doctor	69	22.2 (7.2)
Cost too much	54	17.4 (8.6)
Mistreated by doctor or staff	53	17.0 (8.7)
Had to wait too long	50	16.1 (8.5)
Language problem	25	8.1 *

*Respondents could report more than one problem.

REFERENCES

Aday, L.A. and R. Andersen.
1975 Access to Medical Care, Ann Arbor: Health Administration Press,

Aday, L.A., G. Fleming and R. Andersen.
1984 Access to Medical Care in the U.S.: Who Has It, Who Doesn't, University of Chicago: Pluribus Press, 1984.

Andersen, R.M., A. Giachello, and L.A. Aday.
1986 "Access of Hispanics to Health Care and Cuts in Services: A State-of-the-Art Overview, Public Health Reports 101(3):238-252, May-June.

Carter, O., C. Moy, and F. Trevino.
1985 "Health Services Utilization and Cardiovascular Risk Factors Among Mexican-Americans: HHANES, Southwest United States, 1982-1983. Paper presented at the Annual Meeting of the American Public Health Association, November, 1985, Washington, D.C.

Food Research and Action Center.
1986 CCHIP Questionnaire, Draft 7. Washington, D.C.: FRAC.

Health Resources Administration.
1980 Health of the Disadvantaged, Chart Book II. DHHS Publication No. (HRA) 80-633. U.S. Government Printing Office, Washington, D.C.

Kish, L.
1965 Survey Sampling, John Wiley & Sons, Inc.

Littlefield, C. and C. Stout.
1985 "Migrant Maternity Study", Denver, Colorado: Colorado Department of Health, Migrant Health Program.

Littlefield, C., C. Stout, P. Ramirez-Flores.
1986 Family Planning for Migrant Farmworkers of Mexican Culture: A Framework for Action, Denver, Colorado: Colorado Community Health Network.

Migrant Health Task Force
1986 Report: Meeting at Annual Migrant Health Conference, Minneapolis, Minnesota, April.

Migrant Health Task Force
1987 Report: Meeting at Annual Migrant Health Conference, Padre Island, Texas, April.

National Center for Health Statistics
1981 Visits to Family Planning Clinics: U.S., 1979. Advance Data No. 74, September.

National Center for Health Statistics
1984 Health Indicators for Hispanic, Black, and White Americans, Data from the National Health Survey, Series 10, No. 148, DHHS Publication No. (PHS) 84-1576.

Norusis, M.
1986 SPSS/PC+ for the IBM PC/XT/AT, Chicago: SPSS Inc.

President's Commission for the Study of Ethnical Problems in Medicine and Biomedical and Behavioral Research.
1983 Securing Access to Health Care, A Report on the Ethical Implications of Differences in the Availability of Health Services, Volume One: Report.

Quesada, G. and P. Heller.
1977 "Sociocultural Barriers to Medical Care Among Mexican Americans in Texas: A Summary Report of Research Conducted by the Southwest Medical Sociology Ad Hoc Committee, Medical Care 15(5), May Supplement.

Sakala, C.
1985 "Sociocultural Barriers to Appropriate Health Care for Ethnic Minorities: A Review of Key Issues," Technical Report, Salt Lake City, Utah: Utah Department of Health.

Slesinger, D. and E. Cautley.
1979 "Health Needs of Migrant Workers in Wisconsin, Madison, Wisconsin: Bureau of Community Health, Division of Health.

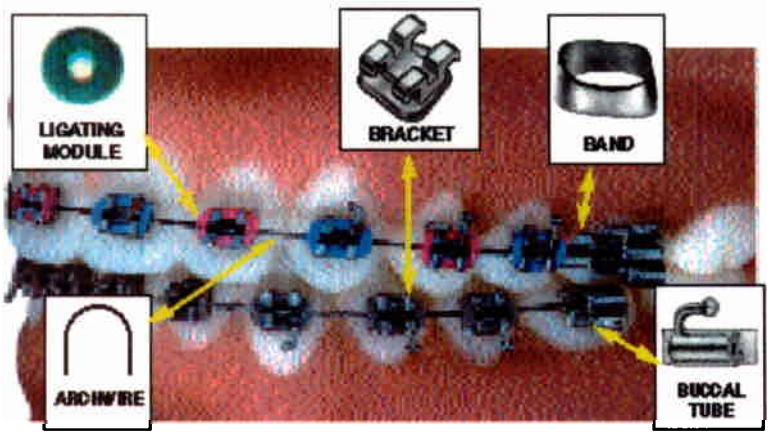
Smith, J.C., C.W. Warren, and N.J. Garcia.
1983 The U.S.-Mexico Border: Contraceptive Use and Maternal Health Care in Perspective, 1979. El Paso: United States-Mexico Border Health Association.

Utahns Against Hunger and Utah Department of Health.
1986 Utah Nutrition Monitoring Project, Salt Lake City, Utah: Utah Department of Health.

Question 2:

Investigators in the Department of Orthodontics at UNC conducted a study to determine the type of orthodontic braces with the lowest amount of friction between the bracket and archwire (high friction works against the desired tooth movement, which is often accomplished by slowly sliding the tooth). Investigators tested two types of bracket (stainless steel and polycrystalline aluminum[1] "invisible" brackets), four types of archwire (stainless steel, cobalt-chromium (Co-Cr), nickel titanium (NiTi), and beta-titanium (β-Ti)), two different bracket slot and archwire sizes (0.018" x 0.025" wires in 0.018" bracket slots and 0.021" x 0.025" wires in 0.022" bracket slots), two surface states (dry and wet using saliva), and two types of friction measurements (kinetic and static[2]). The tests were performed by two different lab technicians, denoted operator 1 (who took measurements in the dry and wet surface states) and operator 2 (who took measurements only in the wet surface state). The outcome of interest to the investigators is the frictional coefficient. The frictional coefficient is the slope of the line obtained from the regression of drawing forces (friction measurements) on the forces applied by the operator using a machine. That is, each frictional coefficient equals the slope of a regression line. Although we do not have the original data used to obtain the frictional coefficients, we do have the frictional coefficients and their variances.

Figure 1: Archwire and brackets



The data are available in the file O:\BIOSLIB\BASICS\2001\ORTHO.SD2, a SAS dataset with 96 observations and 8 variables: BRACKET (steel or "invisible"), FRICTYPE (kinetic or static), SURFACE (dry or wet), OPERATOR (1 or 2), WIRESIZE (0.018" or 0.021"), ARCHWIRE (Steel, Co-Cr, NiTi, or β-Ti), FRICTION (frictional coefficient), and FRICVAR (the variance of the estimated friction coefficient in FRICTION).

Specific questions of the investigator:

1. If a patient requests "invisible" brackets, which archwire type should be used?
2. Which archwire type is superior for stainless steel brackets?
3. Are "invisible" brackets significantly worse (worse=higher friction) than stainless steel brackets?
4. The operators used their own saliva for the measurements made in the wet states. Is there an operator effect within surface?
5. What are the average frictional coefficients for each archwire, bracket, and bracket slot/archwire size combination? Which combination of archwire, bracket, and bracket slot and archwire size results in the lowest amount of friction? Are any other combinations equivalent to the best one?

Possible issues of concern:

1. The investigators are particularly interested in the effects of the bracket type, archwire type, and bracket slot and archwire sizes. Interactions among these effects, including higher-order interactions, may have important scientific meaning to the investigators.
2. The frictional coefficients (outcomes) are slopes of regression lines and do not all have the same variance.
3. The study design is not balanced.
4. There is no replication.

---

[1] Polycrystalline aluminum and several of the wire materials are byproducts of NASA's space research program.

[2] Consider a heavy box on the floor. Static friction is the type of friction that makes it difficult to start pushing the box across the floor. Kinetic friction is the type of friction that acts upon the box while it is moving.

Note: All results should be presented in language the investigators (who are not statisticians) can understand. DO NOT turn in SAS output, though you may wish to create tables based on your SAS output.

1) Fit a fixed effects ANOVA model to the data using reference cell coding. Write out your *final* model (after all fine tuning) and explain the steps you followed to choose that final model. Treat FRICTION as the outcome and BRACKET, FRICTYPE, SURFACE, OPERATOR, WIRESIZE, and ARCHWIRE as potential covariates. (Be sure to include an ANOVA table for your final model.)
2) Using the model you fit in (1), address the specific questions of the investigator. In each case, give your answers in language the investigator can understand.
3) Using the model you fit in (1), explain whether you think the possible issues of concern are important, and explain how you addressed (or decided not to address) those issues in your analysis in (1) and (2).
4) The fixed effects ANOVA model requested in (1) may or may not be the best approach to the data analysis. Comment on whether or not you think fixed effects ANOVA is the most appropriate statistical method in this case. If so, explain why. If not, propose an alternative analysis, and explain why the alternative analysis is superior to your model in (1).

Grading: (1) 6 points, (2) 10 points, (3) 4 points, (4) 5 points

Question 3:

1. Dialysate sodium and ultrafiltration profiling are two methods to reduce symptoms during dialysis (process for which uric acid and urea are removed from circulating blood by means of a dialyzer). The objective of this study was to determine the efficacy of combining these techniques on reducing symptoms in chronic hemodialysis patients. Blood volumes were measured to determine if any observed benefit of profiling could be explained through the combination of dialysate sodium and ultrafiltration profiling. Patients were randomized to a) profiled dialysate sodium and ultrafiltration (profiled treatment) or b) constant dialysate sodium and ultrafiltration (non-profiled, standard treatment).

   The study design is a two-period, two-treatment, two-sequence crossover design with repeated measures, i. e., treatments are given to two groups in two periods following two sequences. Thirty-three patients were randomized. Blood volume data from both standard and profiled treatments were available from 31 patients. Blood volume was recorded at 36-second intervals throughout the dialysis treatment for up to 4 hours.

   Discussion with the scientists led to choosing a model with the following fixed effect sources of variation: treatment, period, sequence, time, quadratic time, and two-way interactions with time and quadratic time (treatment, period, sequence each by time and quadratic time). The scientists are confident that the effect of either treatment is not carried over from one period to another. Therefore you may assume there is no carry-over effect, i. e., no period by treatment effect directly or by time. Experience with similar data makes the assumption of Gaussian errors plausible.

   The data is provided in the SAS file named O:\BIOSLIB\BASICS\2001 Part 2 Exam.SD2 (SAS v6.12, which can be read by later versions automatically). The file contains appropriate labels to properly identify all variables needed for analysis.

a. Provide a very brief descriptive study of the data (no more than one half of a page and 2-3 tables).
b. Carefully state an appropriate linear model for the data, including (mean) response function, covariance matrices, and distributional assumptions. Response vectors, design matrices and parameters should all be well labeled, including the dimensions.
c. Fit the model stated in (b). Provide estimates, standard errors, and tests of hypotheses for all fixed effect parameters and covariance parameters.
d. Reduce the model in (c), as guided by appropriate hypothesis tests and associated analysis. At each step, provide estimates, standard errors, and tests for all fixed effect parameters and covariance parameters. Explicitly identify your final model and why it was chosen.
e. Briefly discuss/explain whether there is a benefit of profiling on blood volume.

points: a) 3 b) 10 c) 5 d) 5 e) 2

Question 4.

Scientific Objective. To examine the association of a patient receiving a health maintenance visit (HMV) with patient, physician, and practice characteristics in the primary care setting.

Scientific background. Because of a strong association between health maintenance visits (HMVs) and cancer screening, knowledge of the predictors of HMV has implications for screening.

Design. A statewide study of cancer screening was conducted in Colorado to determine concordance with the National Cancer Institute's guidelines for screening for breast, cervical, prostate, and skin cancer. Medical records from patients were randomly chosen from primary care practices. The goal was to obtain approximately 22 female patients and 22 male patients from each practice. Whether or not the patient had an HMV in the previous year was recorded.

Participants. A total of 5746 patients aged 42 to 74 years from 132 primary care practices. Additionally, one physician per practice was surveyed; only patients of that physician were selected to be in the sample.

Variables to Consider.

PHYSID	a numeric identifier for physician

All of the following variables are binary (0/1) indicators with a value of 1 assigned for the category described next to the variable name:

HMV	whether the patient had a health maintenance visit in the previous year
LARGE	whether the practice has 3 or more physicians
MDFEM	whether the physician is female
READYMD	whether the physician is ready to change medical practice behaviors
PATFEM	whether the patient is female
NOSMOKER	whether the patient is a non-smoker
AGECAT2	whether the patient is between 50 and 59 years old
AGECAT3	whether the patient is between 60 and 69 years old
AGECAT4	whether the patient is between 70 and 74 years old

Note that since age is rounded to the nearest year the variables AGECAT2, AGECAT3 and AGECAT4 correspond to four mutually exclusive age categories with the reference category being 42-49 years of age.

A listing of the data for one physician is given on the next page.
The full SAS data set is in O:\BIOSLIB\BASICS\2001\HMV.SD2 DAT
Respond to the questions that follow. DO NOT turn in any SAS output, though you may choose to make summaries of the results from your SAS output.

Listing of data for one physician

Obs	PHYSID	HMV	mdfem	large	readymd	nosmoker	patfem	agecat2	agecat3	agecat4
1	15	0	0	1	0	1	1	0	0	1
2	15	1	0	1	0	1	1	0	0	1
3	15	1	0	1	0	1	1	0	1	0
4	15	1	0	1	0	1	1	0	1	0
5	15	0	0	1	0	1	1	0	1	0
6	15	1	0	1	0	1	1	1	0	0
7	15	0	0	1	0	1	1	1	0	0
8	15	0	0	1	0	1	1	1	0	0
9	15	0	0	1	0	0	1	1	0	0
10	15	0	0	1	0	1	1	1	0	0
11	15	0	0	1	0	1	1	1	0	0
12	15	1	0	1	0	1	1	1	0	0
13	15	0	0	1	0	1	1	1	0	0
14	15	1	0	1	0	0	1	0	0	0
15	15	1	0	1	0	1	1	0	0	0
16	15	0	0	1	0	1	1	0	0	0
17	15	1	0	1	0	1	1	0	0	0
18	15	1	0	1	0	1	1	0	0	0
19	15	0	0	1	0	1	1	0	0	0
20	15	0	0	1	0	0	1	0	0	0
21	15	1	0	1	0	1	1	0	0	0
22	15	0	0	1	0	1	1	0	0	0
23	15	1	0	1	0	0	0	0	0	1
24	15	1	0	1	0	1	0	0	1	0
25	15	0	0	1	0	1	0	0	1	0
26	15	0	0	1	0	1	0	0	1	0
27	15	1	0	1	0	1	0	0	1	0
28	15	1	0	1	0	1	0	0	1	0
29	15	0	0	1	0	1	0	0	1	0
30	15	1	0	1	0	1	0	0	1	0
31	15	1	0	1	0	1	0	0	1	0
32	15	0	0	1	0	0	0	1	0	0
33	15	0	0	1	0	0	0	1	0	0
34	15	0	0	1	0	1	0	1	0	0
35	15	0	0	1	0	1	0	1	0	0
36	15	1	0	1	0	1	0	1	0	0
37	15	0	0	1	0	1	0	0	0	0
38	15	0	0	1	0	1	0	0	0	0
39	15	0	0	1	0	1	0	0	0	0
40	15	0	0	1	0	1	0	0	0	0
41	15	0	0	1	0	0	0	0	0	0
42	15	0	0	1	0	0	0	0	0	0
43	15	0	0	1	0	1	0	0	0	0
44	15	0	0	1	0	1	0	0	0	0

1. Provide descriptive statistics for the 5746 patients in the sample.

2. Provide descriptive statistics for the 132 physicians in the study (with respect to variables LARGE, MDFEM and READYMD).

3. Let n_i = the number of patients sampled in the ith practice. Describe the sample distribution of the n_i, i = 1 ,..., 132. Present this graphically (with a histogram or stem-and-leaf plot or box-plot). Report the mean and mode of the sample.

4. Let y_i = the number of patients in the ith practice who had an HMV. Using only the practice and physician predictors (and no interactions) write down a logistic regression model for the binomial response y_i/n_i, i = 1 ...132. Be precise with notation. Assume $Var(Y_i) = n_i\pi_i(1-\pi_i)$, that is to say, no overdispersion. Fit the model using maximum likelihood giving both the estimated regression coefficients and their standard errors.

5. Suppose all clusters have the same size, $n_i = n$ for i = 1,... K. Suppose $Y_{i1}, Y_{i2}, ..., Y_{in}$ are Bernoulli random variables such that $E[Y_{ij}] = \pi_i$ and Corr (Y_{ij}, Y_{ik}) = ρ for $j \neq k$ (exchangeable correlation). Give an expression for Var[Y_i] where $Y_i = \Sigma Y_{ij}$ with the summation over j = 1,...,n. What does this expression reduce to when $\rho = 0$?

6. Refit the logistic regression model in part 4., but this time assume $Var(Y_i) = n_i\pi_i(1-\pi_i)\,\phi$ where ϕ is the overdispersion parameter constant across clusters of possible varying size. What is the estimate of ϕ? Give the estimated regression coefficients and their standard errors. HINT: This is quasi-likelihood in which the estimate of ϕ is given by the square of the "SCALE" parameter in SAS PROC GENMOD obtained with option SCALE=P.

7. Again, using only the practice and physician predictors (and no interactions) write down a logistic model for the binary response, y_{ij} or HMV, of the jth patient in the ith practice. Be precise with your notation. Obtain estimated regression coefficients by applying generalized estimating equations using an independence working correlation structure. Report both model-based (using GENMOD option "MODELSE" on the REPEATED statement) and empirical/robust standard errors. Compare or draw contrast with the analyses in parts. 4 and 6. Repeat the analysis using exchangeable working correlation. Report the estimate of ρ given by SAS. Estimate a crude measure of overdispersion using the formula, $1 + \rho$ (ave(n_i) – 1) where ave(n_i) is the average cluster size.

8. Apply the generalized estimating equations procedure to identify a logistic model to explain the probability that a patient has an HMV as a function of practice, physician and patient characteristics. Give attention to possible two-way interactions. Justify choice of a parsimonious model. Interpret the results providing clear explanations of the direction and magnitude of relevant odds ratios and give their 95% confidence intervals. Base your results on the exchangeable working correlation and empirical/robust standard errors.

Grading: (1) 2 points, (2) 2 points, (3) 2 points, (4) 3 points, (5) 3 points, (6) 3 points, (7) 4 points, (8) 6 points.

Question 5:

A two periods, two treatments crossover clinical trial was conducted to compare a test treatment (T) to placebo (P) for relieving symptoms of a chronic gastrointestinal disorder. In this study, patients initially had evaluation at a screening visit at which the extent of the disorder during the previous 8 weeks was rated as low frequency (1), moderate frequency (2), or high frequency (3); also, age and gender were recorded at this visit. After qualifying for entry to the study, each patient received the randomly assigned treatment (either T or P) for the first period of 14 days of treatment. At the beginning of this period (baseline), the patients rated their symptoms as moderate (1) or severe (2). The primary response variable was their rating of relief at the end of the period as none (0) or adequate (1). After the first period, there was a 28 day washout period with no treatment other than the patients' usual medication. The patient then began the second period of 14 days treatment with the opposite treatment to that during the first period. For this second period, each patient provided a baseline rating of severity of symptoms at the beginning and a response rating of relief at the end. SAS code with a DATALINES statement (data for 2 periods of 260 patients, or 520 records) is available on the BIOS 165 webpage ( ). The data included in the SAS code is comprised of 520 records and 9 variables: patient identification number (PID), sequence (GROUP with P/T for placebo during first period and test treatment during second period, and T/P for test treatment during first period and placebo during second period), PERIOD (1 or 2), treatment (TRT for placebo (P) or test treatment (T)), AGE (in years), SEX (FEMALE or MALE), extent of disorder at screening (SCREEN as (1) for low frequency, (2) for moderate frequency, or (3) for high frequency), severity at baseline (BASE as (1) for moderate, or (2) for severe), and relief at the end of the treatment period (RESPONSE as (0) for none, or (1) for adequate.

3 pts. a. For each treatment group, determine the proportion of patients with complete relief during the first period. Provide a two-sided 0.95 confidence interval for the difference between test treatment and placebo in such proportions. Interpret this result in terms of the comparison between these two treatments for the probability of complete relief. (Hint: use only the records with PERIOD=1).

3 pts. b. For the first period, describe the distributions of age, sex, extent of disorder at screening, and severity at baseline for each treatment group. Apply statistical tests to compare placebo and test treatment for these distributions. Interpret the results. (Hint: use only the records with PERIOD=1).

3 pts. c. Statistically compare the two treatments with respect to relief during the first treatment period in a manner that accounts for any potential influence of GENDER, extent of disorder at screening, and severity at baseline. State assumptions and interpret results. (Hint: use only the records with PERIOD=1).

5 pts. d. For the data from Period 1, apply a statistical model which describes the relationship of the probability of complete relief to the effects of gender, age, extent at screening, severity at baseline, treatment, and any noteworthy interactions among these factors. Discuss assumptions for the model and considerations for evaluating goodness of fit. Provide two-sided 0.95 confidence intervals for the parameters which are relevant to interpreting the factors in this model and discuss their nature. (Hint: use only the records with PERIOD=1).

4 pts. e. For the data from both Period 1 and Period 2, use a statistical test to compare the two treatments for the probability of complete relief in a manner that accounts for the crossover design of the clinical trial. Also provide a two-sided 0.95 confidence interval for a measure of

the association between test versus placebo treatment and complete relief versus not complete relief. Discuss assumptions and interpret results.

3 pts. f. Address the objectives in (e) with conditional logistic regression models which enable evaluation of the potential influence of interactions of treatment with gender and extent at screening. Discuss assumptions and interpret results.

4 pts. g. For the data from both Period 1 and Period 2, apply a statistical model which describes the relationship of the population average probability of complete relief to the effects of gender, age, extent at screening, severity at baseline, period, treatment, and any noteworthy interactions among these factors. Discuss assumptions for the model and considerations for evaluating goodness of fit. Provide two-sided 0.95 confidence intervals for the parameters which are relevant to interpreting the factors in this model and discuss their nature.

List of first 10 records

PID	GROUP	TRT	PERIOD	SCREEN	BASE	RESPONSE	SEX	AGE
1	P/T	P	1	1	2	0	FEMALE	47
1	P/T	T	2	1	2	1	FEMALE	47
2	T/P	T	1	3	2	0	FEMALE	69
2	T/P	P	2	3	2	0	FEMALE	69
3	P/T	P	1	1	1	1	MALE	78
3	P/T	T	2	1	1	1	MALE	78
4	T/P	T	1	1	1	1	FEMALE	60
4	T/P	P	2	1	2	0	FEMALE	60
5	P/T	P	1	1	1	1	FEMALE	70
5	P/T	T	2	1	1	1	FEMALE	70

Variables Label

#	Variable	Type	Len	Pos	Format	Informat	Label
2	AGE	Num	8	0	BEST12.	12.	Patient Age at Each Period
6	BASE	Num	8	24			Severity at Period Baseline
1	GROUP	Char	17	80			Treatment Sequence
4	PERIOD	Num	8	8			Study Period
8	PID	Num	8	40			Patient ID
7	RESPONSE	Num	8	32			Relief After Treatment
5	SCREEN	Num	8	16			Extent at Screening
3	SEX	Char	12	97			Gender
9	TRT	Char	1	109			Treatment

Question 6:

The data set shown below was obtained by asking a random sample of 637 voters to rank three presidential candidates, here labelled A, B and C, in order of preference. Partial ranking and ties were not allowed. (Partial ranking occurs when only two of the three choices are ranked. A tie occurs when two or more choices are ranked equal.) The response from each subject can be represented by a permutation of the letters A, B and C, The data were as follows:

Ordering	Count
ABC	23
ACB	210
BAC	8
BCA	81
CAB	111
CBA	204
Total	637

For example, 23 subjects listed A as their first choice, B as their second choice and C as their third choice.

Assume that the 637 responses are independent and identically distributed random variables, each having a multinomial distribution with index 1 and probability vector $\pi = \{\pi_{ijk}, (i,j,k) \in \Gamma\}$, where $\pi_{ijk} = pr(A \text{ has rank } i, B \text{ has rank } j, C \text{ has rank } k)$, and Γ is the set of all 6 permutations of the numbers $\{1,2,3\}$.

1. (4 points) Define three factors, A, B and C, each with three levels, such that A at level 3 means that candidate A was ranked in third position, and so on for the remaining factors. For example, for the 111 subjects who responded CAB: $A = 2, B = 3$ and $C = 1$. Clearly, only 6 combinations of factor levels can occur and they correspond to the 6 permutations of the numbers {1, 2, 3}.

   Fit the log-linear models 1 and $1 + A + B + C$. For coding purposes use level 1 as the reference cell (standard GLIM restrictions). Report the deviances, degrees of freedom, parameter estimates and standard error estimates. What is the rank of the latter model matrix? Explain in detail why C can be dropped without affecting the fit.

2. (5 points) Consider the log-linear model $A * B$. Provide a precise and meaningful interpretation of the parameter(s) corresponding to $A.B$ with reference to the achievability of a unique ranking of the three candidates under this model. Within this model, test the hypothesis of null $A.B$ interactions against an unrestriceted alternative. Use a likelihood-ratio test at the 5% level. Interpret this hypothesis.

3. (5 points) Define linear contrasts for each of the three factors such that C_L takes values $-1, 0, 1$ and C_Q takes values $1, -2, 1$ for the three levels of C. Fit the model

$$1 + A_L + B_L + C_L.$$

Which candidate has the smallest coefficient? Provide precise and meaningful interpretations of these coefficients.

4. (5 points) Add the terms

$$A_Q + B_Q + C_Q$$

to the previous model. Which candidate has the largest quadratic coefficient? Interpret the sizes of the quadratic coefficients in terms of heterogeniety among voters and negative voting.

5. (5 points) Examine the two-way table of total votes indexed by candidate and rank. Compute the fitted values for this table under the quadratic model just fitted. Compare the two tables. If any patterns emerge explain exactly why they occur.

6. (1 point) Which candidate has the most first-place votes? Which candidate is least disliked? Which candidate ought to be declared the winner according to these data.