

STATISTICS QUALIFYING EXAM - TAKE HOME
January, 2007

Directions: The answer to each problem should be presented as summary. It should be word-processed and double spaced. An appendix is allowed for each problem but will be examined only at the discretion of the graders. The better constructed your appendix, the more likely it is to get examined.

You may not consult any other person when working on this exam or discuss your exam with anyone else regardless of whether or not the person is taking the exam. You may use only your class notes for the exam. You may use any available books or web resources for the exam.

1. Four factors are thought to possibly influence the taste of a soft-drink beverage: type of sweetner (A), ratio of syrup to water (B), carbonation level (C), and temperature (D). Each factor can be run at two levels. At each run in the design, samples of the beverage are given to a test panel consisting of 20 people. Each tester assigns point score from 1 to 10 to the beverage. Total score is the response variable, and the objective is to find a formulation that maximizes total score. Two replicates of this design are run, and the results are shown below. Analyze the data and draw conclusions.

Treatment combination	Replicate		Treatment combination	Replicate	
	I	II		I	II
(1)	190	193	d	198	195
a	174	178	ad	172	176
b	181	185	bd	187	183
ab	183	180	abd	185	186
c	177	178	cd	199	190
ac	181	180	acd	179	175
bc	188	182	bcd	187	184
abc	173	170	abcd	180	180

2. Christopher Bingham collected data from the American Almanac of 1974 with the purpose to study fuel consumption in terms of other variables. The data set consists of 9 columns and each row corresponds to the 48 contiguous states. The columns in the data set from left to right are:

- STATE.
- AREA (sq. mi.) .
- 1971 population in thousands [POP].
- 1972 motor fuel tax rate, in cents per gallon [TAX].
- 1971 thousands of licensed drivers [NLIC].
- 1972 per capita income in thousands of dollars [INC].

- 1971 thousands of miles of federal-aid primary highways [ROAD].
- DRIVERS= nlic/pop = proportion of population with driver's licences
- FUEL = fuel/pop=gallons of fuel consumed per person.

The interest is to develop some model or models to relate fuel consumption to other variables. Based on your statistical analysis of these data provide a report that:

- (a) describes the methodology you used to build a model or models, including relevant output.
- (b) Identifies those variables that are most relevant to explain the response and interpret its meaning in the context of your model. Try to support your analyses with simple graphs or plots.
- (c) Evaluates the potential for multicollinearity problems.
- (d) Includes relevant summary tables of the important predictor variables and other features such as coefficients, standard errors or p-values. Give appropriate interpretation of these tables.
- (e) Make an assessment of your model(s) and particularly identify outliers/leverage points and if any transformations are required.
- (f) Write a general conclusion of your analysis, including any potential limitations of your model(s) and possible extensions to improve on your results.
- (g) A particular interest here is to assess whether TAX rate is important in explaining the part of FUEL consumption not explained by DLIC. That is, we are interested in understanding whether TAX will be an important predictor of FUEL consumption when added to a model that already includes DRIVERS. Quantify the unique effect of adding TAX on FUEL to a model that already includes DRIVERS. Justify your answer.

The data are available in <http://www.math.unm.edu/~ghuerta/fuel.txt> as a raw text file.