Graduate Statistics Fall, 1998
Final Exam

There are four major problems (A through D) on this exam. Each sub-question is worth 3 points.

## PROBLEM A

The following question is loosely based on a study reported by Le Clerc and Little (1997) in the *Journal of Marketing Research*. Unlike a number of other recent articles in this journal, their paper correctly interpreted the coefficient for a component term of an interaction. In fact, Le Clerc and Little used the interaction model to ask a sophisticated question, as will you in the following questions.

Random shoppers at a shopping mall were intercepted and invited to participate in a study about advertising. Each shopper was shown one of three ads: Celebrity—the fictitious FOOBAR brand is shown being used by a famous person; Comparison—the FOOBAR brand is compared to other brands of the same product; Family—the FOOBAR brand is shown being used in a stereotypical family setting. Shoppers then rate their attitude toward the FOOBAR brand and answer a few questions.

The following variables are available (note: they have unique first letters which you may use in your answers instead of the full variable names). Use these variables and others that you construct with them (be sure to define them!) to specify MODEL A/C comparisons that could be used to answer the research questions below. Specify PA-PC and the null hypothesis.

| | |
|---|---|
| TYPE of Ad: | Celebrity, Comparison, Family |
| ATTITUDE: | Attitude toward a particular brand (we will call it the FOOBAR brand) on a 1 to 10 scale (with higher scores being more favorable towards the brand) |
| BRANDS: | Number of different brands of this product type purchased in the last year. |
| YEARS: | Age |

1. Do older people become “set in their ways” and buy fewer brands of this product?

2. A company is particularly interested in the typical number of brands purchased by people who are 25 years old. In particular, the company would like to know whether the number of brands purchased by people in that age group is greater than 2. Assuming that age is a useful predictor of the number of brands purchased, provide the most powerful test of the company’s question.

3. Do age and the number of brands purchased predict attitude towards the FOOBAR brand?

4. Does the effect of age on attitude toward the FOOBAR brand depend on the number of brands purchased?

5. Assuming the answer to the previous question is yes, is there an effect of age on attitude towards the FOOBAR brand for brand-loyal customers? Someone is brand-loyal if they only purchase one brand in the previous year.

6. Do the three types of ads produce different mean attitudes toward the FOOBAR brand?

7. In particular, does the comparison ad produce a lower attitude than the types of ads showing the product being used by either a celebrity or a family?

8. What is the answer to the previous question when controlling for age?

9. Returning to the question of whether the three types of ads produce different mean attitudes, is the answer to this question the same for older as for younger shoppers?

10. Does the answer to the prior question depend on the number of brands the shopper has purchased in the previous year?

Problem B

Two different experimenters test the exact same hypothesis concerning food deprivation and learning. Rats are randomly assigned to one of three groups: no deprivation, deprived for three hours, deprived for six hours. They then are asked to run a maze on which they have been previously conditioned, to find a food reward at the end. The time it takes them to run the maze is the dependent variable.

The first experimenter has 18 rats, so he has 6 in each of the three conditions. The second has 36 rats, so he has 12 in each of the three conditions.

Both experimenters find exactly the same means. In the no deprivation condition, the mean running time is 20 seconds; with 3 hours deprivation, the mean running time is 16 seconds; and with 6 hours deprivation, the mean running time is 12 seconds.

Assume both test the same single degree of freedom contrast, asking whether mean running time in the no deprivation condition differs from that in the six hour condition. The first reports an F* of 8.44 (PRE = .36). The second reports an F* of 7.74 (PRE = .19).

1. What are the degrees of freedom for each of these F*'s?

2. What are the SSR's associated with the tested contrast in each experiment?

3. Do you think it likely that the two groups of rats, those run by the first experimenter and those run by the second, were sampled from the same population of rats? Would one be justified in doing a single analysis for all 54 rats, ignoring which experiment they were in? Why or why not?

4. When these two researchers meet at a convention, they discover that the first one used lab rats of a particular strain while the second one used wild rats. The first researcher decides to replicate the study using wild rats and the first sample size of 18. Approximately what are the chances that a significant effect for the contrast of interest will be found.

Problem C

The regressions on the following pages resulted from continuing analyses of the USNEWS dataset, containing statistics on colleges in the United States. We are still trying to understand factors that are associated with higher or lower graduation rates in these colleges. The following variables are used in the analysis:

| | |
|---|---|
| GR | Graduate rate (0 - 100) |
| SIZE | Number of fulltime students/1000 |
| SAT | Combined SAT scores of admitted class (400-1600) |
| PUBPRI | Public or private institution (1=Public; -1=Private) |

Some additional variables were also used that are functions of the above variables. These are defined as:

| | |
|---|---|
| I1 | SIZE*SAT |
| SIZED | SIZE minus its mean (3.89) |
| SATD | SAT minus its mean (976.29) |
| I1D | SIZED*SATD |

Based on the following output, answer the following questions:

1. Assuming that SAT and SIZE are additive in predicting GR, do larger schools have lower graduatation rates even when we equate them on the SAT scores of entering students? (Provide SSE(A), SSE(C), PRE, F*, and N-PA in addition to your answer.)

2. Is it the case that the effect of SIZE on GR depends on SAT? (Provide SSE(A), SSE(C), PRE, F*, and N-PA in addition to your answer.)

3. Assuming that your answer to question B was “yes” (that’s a hint!), write a couple of sentences that describe how the effect of SIZE on GR changes as SAT gets larger.

4. According to the model that allows the effect of SIZE on GR to depend on SAT, what is the simple slope for SIZE in a school where the average SAT score of entering students is 1000?

5. Provide a short interpretation of the slope for SAT in Model 2 (i.e., .073881).

6. Assuming that the effect of SIZE on GR does depend on SAT, is there a significant effect of SIZE on GR for schools at the average value of SAT. (Provide SSE(A), SSE(C), PRE, F*, and N-PA in addition to your answer.)

7. Assuming additivity, does SIZE continue to affect GR if we look within public and private institutions (still controlling for SAT)? (Provide SSE(A), SSE(C), PRE, F*, and N-PA in addition to your answer.)

8. According to model 5, which allows the effect of SIZE on GR to depend on SAT, write out the intercept and slope for the SIZE : GR simple relationship for public institutions.

Model: MODEL1
Dependent Variable: GR

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 2 | 89066.99525 | 44533.49763 | 203.576 | 0.0001 |
| Error | 715 | 156410.25266 | 218.75560 | | |
| C Total | 717 | 245477.24791 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 14.79039 | R-square | 0.3628 |
| Dep Mean | 62.40669 | Adj R-sq | 0.3610 |
| C.V. | 23.70001 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -22.984750 | 4.43821316 | -5.179 | 0.0001 |
| SIZE | 1 | -0.672714 | 0.11474254 | -5.863 | 0.0001 |
| SAT | 1 | 0.090147 | 0.00454044 | 19.854 | 0.0001 |

| Variable | DF | Type II SS | Squared Partial Corr Type II | Tolerance |
|---|---|---|---|---|
| INTERCEP | 1 | 5867.085866 | . | . |
| SIZE | 1 | 7519.194397 | 0.04586848 | 0.98673829 |
| SAT | 1 | 86231 | 0.35538567 | 0.98673829 |

Model: MODEL2
Dependent Variable: GR

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 3 | 93315.20102 | 31105.06701 | 145.956 | 0.0001 |
| Error | 714 | 152162.04689 | 213.11211 | | |
| C Total | 717 | 245477.24791 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 14.59836 | R-square | 0.3801 |
| Dep Mean | 62.40669 | Adj R-sq | 0.3775 |
| C.V. | 23.39230 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -6.716281 | 5.69793384 | -1.179 | 0.2389 |
| SIZE | 1 | -5.811577 | 1.15653980 | -5.025 | 0.0001 |
| SAT | 1 | 0.073881 | 0.00577555 | 12.792 | 0.0001 |
| I1 | 1 | 0.005070 | 0.00113566 | 4.465 | 0.0001 |

| Variable | DF | Type II SS | Squared Partial Corr Type II | Tolerance |
|---|---|---|---|---|
| INTERCEP | 1 | 296.094944 | . | . |
| SIZE | 1 | 5381.147527 | 0.03415665 | 0.00946191 |
| SAT | 1 | 34873 | 0.18645025 | 0.59410235 |
| I1 | 1 | 4248.205766 | 0.02716066 | 0.00923602 |

Model: MODEL3
Dependent Variable: GR

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 3 | 93315.20102 | 31105.06701 | 145.956 | 0.0001 |
| Error | 714 | 152162.04689 | 213.11211 | | |
| C Total | 717 | 245477.24791 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 14.59836 | R-square | 0.3801 |
| Dep Mean | 62.40669 | Adj R-sq | 0.3775 |
| C.V. | 23.39230 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 62.062296 | 0.55028146 | 112.783 | 0.0001 |
| SIZED | 1 | -0.861320 | 0.12087461 | -7.126 | 0.0001 |
| SATD | 1 | 0.093605 | 0.00454793 | 20.582 | 0.0001 |
| I1D | 1 | 0.005070 | 0.00113566 | 4.465 | 0.0001 |

| Variable | DF | Type II SS | Squared Partial Corr Type II | Tolerance |
|---|---|---|---|---|
| INTERCEP | 1 | 2710779 | . | . |
| SIZED | 1 | 10821 | 0.06639328 | 0.86622282 |
| SATD | 1 | 90277 | 0.37237093 | 0.95812162 |
| I1D | 1 | 4248.205766 | 0.02716066 | 0.86488738 |

Model: MODEL4
Dependent Variable: GR

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 3 | 105250.33012 | 35083.44337 | 178.636 | 0.0001 |
| Error | 714 | 140226.91779 | 196.39624 | | |
| C Total | 717 | 245477.24791 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 14.01414 | R-square | 0.4288 |
| Dep Mean | 62.40669 | Adj R-sq | 0.4264 |
| C.V. | 22.45616 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -17.237236 | 4.25268027 | -4.053 | 0.0001 |
| SIZE | 1 | 0.187092 | 0.14419317 | 1.298 | 0.1949 |
| SAT | 1 | 0.078409 | 0.00449226 | 17.454 | 0.0001 |
| PUBPRI | 1 | -6.794032 | 0.74844553 | -9.078 | 0.0001 |

| Variable | DF | Type II SS | Squared Partial Corr Type II | Tolerance |
|---|---|---|---|---|
| INTERCEP | 1 | 3226.583826 | . | . |
| SIZE | 1 | 330.639120 | 0.00235234 | 0.56096465 |
| SAT | 1 | 59833 | 0.29907461 | 0.90498682 |
| PUBPRI | 1 | 16183 | 0.10346723 | 0.55566881 |

Model: MODEL5
Dependent Variable: GR

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 106353.07883 | 26588.26971 | 136.263 | 0.0001 |
| Error | 713 | 139124.16908 | 195.12506 | | |
| C Total | 717 | 245477.24791 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 13.96872 | R-square | 0.4333 |
| Dep Mean | 62.40669 | Adj R-sq | 0.4301 |
| C.V. | 22.38337 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > \|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -9.057155 | 5.45969156 | -1.659 | 0.0976 |
| SIZE | 1 | -2.585105 | 1.17494158 | -2.200 | 0.0281 |
| SAT | 1 | 0.070652 | 0.00554054 | 12.752 | 0.0001 |
| PUBPRI | 1 | -6.315881 | 0.77265747 | -8.174 | 0.0001 |
| I1 | 1 | 0.002676 | 0.00112548 | 2.377 | 0.0177 |

| Variable | DF | Type II SS | Squared Partial Corr Type II | Tolerance |
|---|---|---|---|---|
| INTERCEP | 1 | 536.982841 | . | . |
| SIZE | 1 | 944.576283 | 0.00674366 | 0.00839406 |
| SAT | 1 | 31729 | 0.18571012 | 0.59108293 |
| PUBPRI | 1 | 13038 | 0.08568416 | 0.51801493 |
| I1 | 1 | 1102.748713 | 0.00786403 | 0.00861016 |

PROBLEM D

A version of this question is submitted to the statistics newsgroup sci.stat.consult about every other week. We've constructed this one, but it is typical of many such questions. It is important that you be able to answer this question if someone (perhaps even your mentor) asks you this question.

"I regressed first-year undergraduate GPA on the student's entering total SAT score and high-school RANK. Not surprisingly, the coefficients for both SAT and RANK were hugely significant. Then I wanted to see if there was an interaction so I included SAT*RANK in the regression. When I did that the whopping effects for SAT and RANK completely disappeared! What happened?"

Write a succinct note to this person explaining what happened to those "whopping effects."