Answer Key

Sociology 593
Exam 2
March 26, 1999

I. True-False. (20 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A researcher regresses Income on Education. She does NOT include any dummy variables or interaction terms. One implication of this model is that, if it is true, the mean income for blacks will be the same as the mean income for whites. F. If mean levels of education differ across races, mean income will differ too.

2. A researcher believes that job experience has a stronger effect on earnings for men than it does women. If she is correct, this means that the effect of job experience on earnings is non-linear. F. Effects are non-additive; the effect of job experience on earnings depends on a 3rd variable, gender.

3. A researcher has included several extraneous variables in her model. The larger her sample, the less problematic this will be. True. Extraneous vars lead to bigger standard errors. Larger sample sizes help to reduce standard errors.

4. A researcher hypothesizes that IQ positively affects the political liberalism of women but negatively affects the political liberalism of men. She gets

$\beta_{IQ} = 6$

$\beta_{Male} = 0$

$\beta_{IQ*Male} = -4$

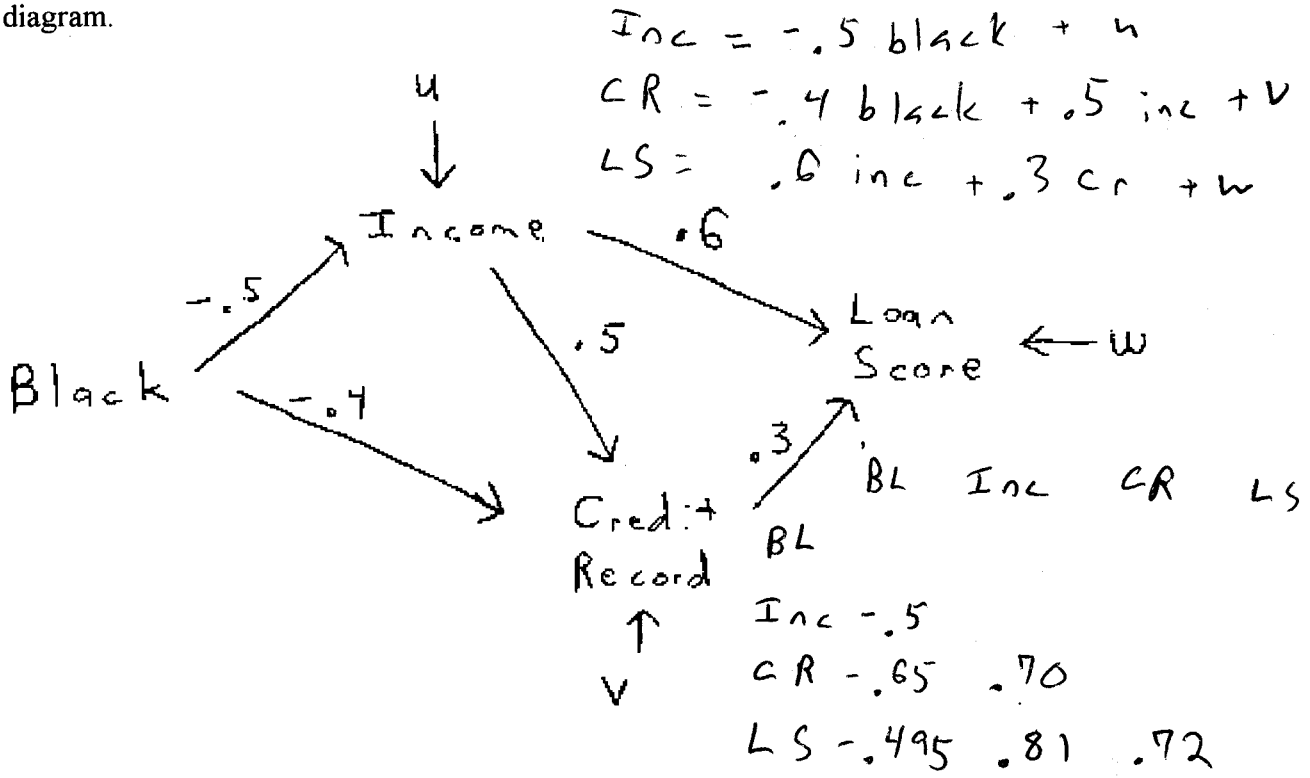
Male = 1 if male, 0 if female. The standard error for the interaction term is 1.0. N = 2,000. The evidence supports the researcher's hypothesis.

F. For women, $E(Liberalism) = \alpha + 6 * IQ$

For men $E(Lib) = \alpha + 1(6-4) * IQ$

So, while the effect of IQ is smaller for men than women, it is still positive.

II. *Path Analysis/Model specification.* (30 points). A researcher is interested in home mortgage lending. She has measures of loan applicant's race, income, credit record (where the higher the score, the better the person's record is) and the Loan Score assigned by the lender (where the higher the score assigned, the stronger the lender thought the application was). All her variables are in standardized form. The hypothesized value of each path is included in the diagram.



a. Write out the structural equation for each endogenous variable.

b. Determine the complete correlation matrix. (Remember, variables are standardized. You can use either normal equations or Sewell Wright, but you might want to use both as a double-check.)

c. Decompose the correlation between Income and Loan Score into

- Correlation due to direct effects .60
- Correlation due to indirect effects .15
- Correlation due to common causes .06

d. Many studies have shown that blacks fair less well in the home mortgage market than would seem justified based on their income alone. That is, even when a black and a white have identical incomes, the black is less likely than the white to get the loan. According to the above model, why does this occur? Would you interpret this model as saying there is racial discrimination against blacks? Why or why not? Blacks have worse credit records than do comparable whites. The model seems to imply discrimination is not a factor (unless discrimination is somehow responsible for blacks having worse credit records.)

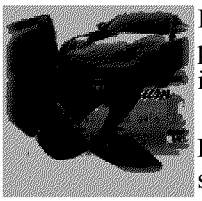
bankrate.com℠: managing your credit: discrimination and credit scoring

Main Story: A credit score can make - or break - a would-be borrower

Table: Credit scoring profiles -- Some examples

Everybody needs an equal chance to borrow -- but do they get it?

By Michael D. Larson
bankrate.com



It's revolutionary. It speeds up the lending process. It allows more people to buy homes, and it makes investors, borrowers and banks happy.

But if all these expert claims about credit scoring are true, why do complaints persist about discrimination against minorities and low-income borrowers? That's a question puzzling just about everyone in the mortgage industry, from Fair, Isaac & Co., which makes the most widely-used scoring systems, to Fannie Mae and Freddie Mac, the government-sponsored corporations that buy mortgages from lenders for resale to Wall Street investors.

In this story:

When lenders become more involved

Equal pain

Where does it begin?
Some say the discrimination begins in the scoring models, which are based on data gathered over decades. Since rampant discrimination occurred during a portion of that time, they argue that bias appears in the scores and makes certain people less likely to be eligible for the automatic approval that comes with a good score.

Others say the problem erupts when human beings are called on to make lending decisions about applicants whose scores fail to provide a definitive approval.

Credit scores are used to speed the lending process by providing a standardized evaluation of most mortgage, credit card or other loan applicants that come along. If the person's score is high enough, the loan will go through almost automatically.

When lenders become more involved
Lenders have to take extra steps to approve loans for people whose scores are less impressive, however.

Rich Riese, project manager for compliance policy at the Treasury Department's Office of Thrift Supervision, says concerns about biased lenders have developed from reports from shoppers and testing groups collecting information on face-to-face lender negotiations. He says those studies indicate discrimination can arise when human decision-makers become involved.

"The solution is to get all

Paul Mondor, senior director of regulatory affairs at the Mortgage Bankers Association of America, says, "What happens is that (credit

that are involved in this to get together and talk this through."

scoring) makes it easy and quick to approve certain people and it refers other people to human underwriters, who have to labor long and hard.

"Two things are going to result: The pricing on that second loan is going to be higher, and they're going to wait longer for that loan."

Equal pain

So-called "disparate impact" rules say lending systems can't include a variable that hurts members of one group more than others. But officials say it's difficult to show evidence of this type of violation, or even figure out if it's happening. And a provision that allows banks to include a variable, if there's a "business necessity" for doing so, further complicates the issue.

On the other hand, Riese notes that credit scoring makes gathering information on discrimination easier because it creates better electronic records. And the people who built the systems maintain that credit scores help level the playing field rather than create problems.

"Credit scoring really is an objective, colorblind means of assessing the information that's pertinent in making a lending decision," says David Shellenberger, product manager of credit bureau products for Fair, Isaac. "We're only using information on a credit file; things such as gender, race, religious background, those things just aren't there."

Still, that's only one side of the argument.

"Fair, Isaac and Fannie and Freddie just take this position that, well it's not discriminatory," Mondor says. "But the Department of Housing and Urban Development, Department of Justice, and private nonprofit community access groups are all saying, 'We have grave concerns, and we think there are discriminatory tendencies.'"

"The solution is to get all the players that are involved in this to get together and talk this through."

Related information:
More personal finance news
Definitions: Banking terms

--Posted: Dec. 4, 1998

III. *Short answer.* Answer *two* of the following three questions. (25 points each; up to 10 points extra credit if you do all 3).

1. Draw a scatterplot that illustrates each of the following. Offer a hypothetical example of when you might observe such a relationship between variables. What OLS assumptions, if any (e.g., additivity, linearity), would be violated if you simply regressed Y on X? What harm would result, e.g. would values be over-estimated, under-estimated, or what? Indicate the model you think should be estimated, e.g. $E(Y) = \alpha + \beta_1 X + \beta_2 X^2$.

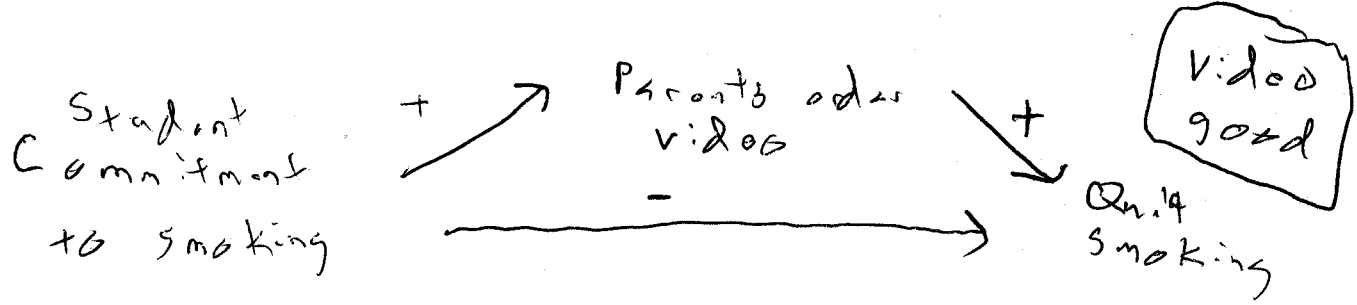
a. The effect of X on Y is positive for small values of X but then switches to being negative. Polynomial, or maybe piecewise

b. The effect of X on Y is the same for both whites and blacks, but the intercepts are different in the two populations.

c. Each one unit increase in X produces a 5% increase in Y. Growth model; tck. LN(Y)

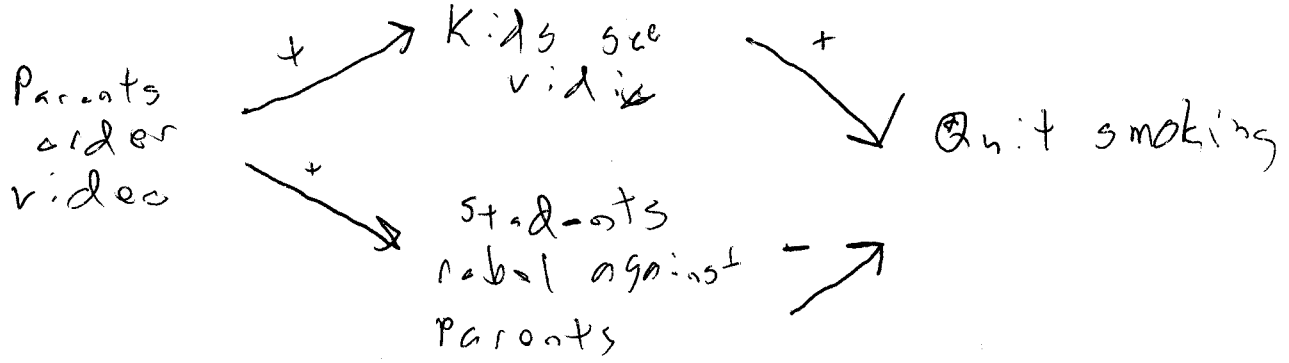
2. The American Cancer Institute is concerned about the recent rise in teenage smoking. It has come up with an anti-smoking video that parents can order for their smoking teenagers. Much to its surprise, its research shows that teen smokers who see the video are less likely to quit than teen smokers who do not. Some critics within the Institute maintain that the video is doing more harm than good and should be abandoned, while others continue to insist that the video is serving its purpose. The Institute has hired you, a professionally trained social scientist, to give it insight on why these relationships exist. Drawing on your knowledge of the logic of causal order, present different models that could account for the observed relationships. Indicate what implications the different models have for what should be done about the video. To be fair, you will want to present one or more models that suggest that the video discourages smoking, one or more models which imply the video encourages smoking, and one or two models which suggest that the video is not achieving what the Institute wants but the problems are correctable (i.e. you don't have to completely scrap the video to solve the problem). When presenting your answer, keep in mind that Institute staff do not know very much about the logic of causal order, so you will have to make things very clear for them.

3. A researcher is interested in how the effect of a message differs by characteristics of the speaker. College student subjects are randomly assigned to read passages whose content ranges from very conservative (MSGLIB = -100) to very liberal (MSGLIB = 100). Half the students are told that the author of the passage is a first year college student (PROF = 0), while the other half are told that the author is a distinguished Nobel Prize winning professor (PROF = 1). Students then complete a questionnaire that measures their own liberalism and conservatism (LIBRLISM = -100 if very conservative, 100 if very liberal). The researcher then runs the following regressions.



Parents of the most hard-core smokers are the most likely to order the video. The video helps somewhat, but the most hardcore are still less likely to quit. There is a selection problem, in that those who see the video are different than those who do not

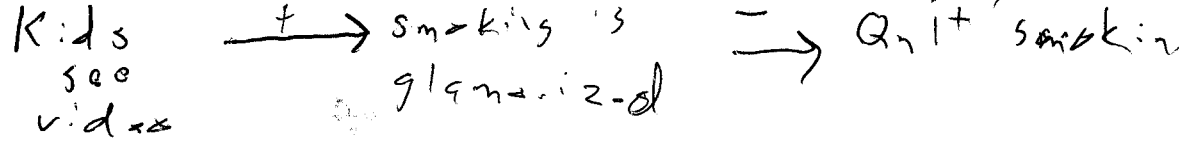
Problems correctible



While the video may be good, kids may rebel against pressure from their parents. Perhaps the Inst could figure out ways for smokers to see the video without the "stigma" of it coming from parents attached.

Video bad

The video may have the unintended effect of glamorizing smoking or making it appear common & acceptable.



Indicate whether there appear to be statistically significant differences in the determinants of liberalism between those who thought the author was a college student and those who thought the author was a professor. If so, tell whether these differences are limited to differences in the intercepts, or whether the effect of the passages differs between the two types of speakers. If differences are found, be specific as to what they are, e.g. how much greater (or weaker) is the effect of messages on beliefs when the author is a professor rather than a student. Briefly discuss the substantive implications of what you think is the best model. Be sure to indicate how the printout supports your arguments.

Go with model 3
when a student is speaking, the effect of message on beliefs is .38.
When the prof is speaking, the effect is almost twice as great, .380 + .313 = .693

```
Compute profmsg = prof * msglib
```

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
LIBRLISM	23.3895	20.1731	200
MSGLIB	44.8919	28.0179	200
PROF	.5000	.5013	200
PROFMSG	22.4459	29.9809	200

Correlations

		LIBRLISM	MSGLIB	PROF	PROFMSG
Pearson Correlation	LIBRLISM	1.000	.746	.429	.772
	MSGLIB	.746	1.000	.000	.467
	PROF	.429	.000	1.000	.751
	PROFMSG	.772	.467	.751	1.000
Sig. (1-tailed)	LIBRLISM	.	.000	.000	.000
	MSGLIB	.000	.	.500	.000
	PROF	.000	.500	.	.000
	PROFMSG	.000	.000	.000	.
N	LIBRLISM	200	200	200	200
	MSGLIB	200	200	200	200
	PROF	200	200	200	200
	PROFMSG	200	200	200	200

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	MSGLIB[a]	.	Enter
2	PROF[a]	.	Enter
3	PROFMSG[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: LIBRLISM