

Soc 63993, Advanced Social Statistics II
Homework No. 1
Review of Multiple Regression

I. The attached table is from an article in Journal for the Scientific Study of Religion, 1990, Vol.29(3):297-314, *Religious Practice: A Human Capital Approach*, written by Laurence Iannaccone. (Table appears on p. 304.) Read the table carefully and answer the following questions.

1. When the dependent variable is CONTRIBUTE, the unstandardized coefficient of INCOME is 9.025. What does this mean?
2. When the dependent variable is CONTRIBUTE, the significance level of INCOME is $p \leq .001$. What does this mean? Which type error does it test? Write out the null and alternative hypotheses that are being tested by the T statistic.
3. Suppose the researcher believes that older people tend to attend fewer masses than do younger people. Do these results support her?
4. What is the standard error of the coefficient for MARSAME in the CONTRIBUTE equation?
5. For the CONTRIBUTE equation, test whether or not $R^2 = 0$ (i.e. compute the F value and degrees of freedom).
6. When a male and female have identical values on all other IVs, which one will be likely to attend mass more (and by how much)?
7. Suppose you somehow managed to get a score of 0 on all the IVs — how many masses a year would you be expected to attend?
8. Among the three multiple regression equations, three independent variables are not statistically significant in any one of them (HEDUC, NBHD, RAISECA). Why do you think the researchers have still included them in the equations?
9. In the ATTEND equation, BELIEF has the largest partial regression coefficient. Can you therefore say that BELIEF is the most important determinant of Church attendance? Why or why not? What other sorts of information might aid you in determining which variable has the strongest influence?

TABLE 1

PARTICIPATION REGRESSIONS:1974 CATHOLIC SURVEY						
Variable	CONTRIBUTE		ATTEND		RATIO (A/C)	
	coefficient	t-stat	coefficient	t-stat	coefficient	t-stat
MARSAME	70.984***	4.60	11.836***	5.25	-0.105**	-2.64
RLGINSTR	3.313*	2.14	0.509*	2.25	-0.000149	-0.04
PCHURCH	0.512	1.63	0.219***	4.80	0.000909	1.19
NOINCOME	172.257***	3.58	6.199	0.88	-0.425***	-3.30
INCOME	9.025***	9.73	0.019	0.14	-0.034***	-4.53
HEDUC	4.124	1.75	0.231	0.67	-0.007	-1.35
AGE	3.702***	7.66	0.316***	4.47	-0.018*	-2.47
SEX	-0.912	-0.07	7.232***	4.13	0.103***	3.70
NONWHITE	-36.259	-1.83	-0.723	-0.25	0.252***	5.34
NKIDS	4.646	1.17	-0.020	-0.03	-0.023*	-2.38
BELIEF	35.256**	2.96	12.591***	7.23	0.039	1.40
NBHD	-25.373	-0.95	-2.725	-0.69	-0.035	-0.56
RAISECA	-5.044	-0.12	-1.068	-0.18	-0.036	-0.37
PCATH	-64.239	-1.62	-11.243**	-1.94	0.025	0.26
(CONSTANT)	-194.256	-4.02	4.122	2.58	1.276	7.64
R-squared	.39		.28		.31	
Cases	555		555		456	

*p ≤ .05 **p ≤ .01 ***p ≤ .001

NOTES:

Coefficients: Unstandardized regression coefficients.

Source: N.O.R.C. American Catholic Survey, 1974.

Sample: All married respondents.

Variable definitions:

AGE = respondent's age.

ATTEND = yearly number of masses attended.

BELIEF = 9-item additive scale of respondent's strength of religious belief.

CONTRIB = yearly contributions to church (excluding Catholic school tuition and contributions).

HEDUC = years of education of family head.

INCOME = yearly income (thousands).

MARSAME = coded 1 if respondent and spouse of same religion.

NBHD = fraction of Catholic neighbors when growing up.

NKIDS = number of preschool or school-age children.

NOINCOME = dummy (1 if income not reported, 0 otherwise).

NONWHITE = dummy (1 if respondent is nonwhite, 0 otherwise).

PCATH = dummy (1 if either parent Catholic, 0 otherwise).

PCHURCH = mean of parents' yearly mass attendance.

RAISECA = dummy (1 if respondent was raised a Catholic, 0 otherwise).

RATIO = time intensity of religious participation - ATTEND/CONTRIB.

RLGINSTR = respondent's religious instruction scale score.

SEX = sex of respondent (1 if female, 0 if male).

II. Download the file *sphrd.dta* from the course web page. As explained in the handout on using Stata for OLS regression, this data set was created using Stata's `corr2data` command based on results published in the 1985 ASR paper, "Ability grouping and contextual determinants of educational expectations in Israel." In that piece, Shavit and Williams examined the effect of ethnicity and other variables on the achievement of Israeli school children. There are two main ethnic groups in Israel: the Ashkenazim - of European birth or extraction - and the Sephardim, most of whose families immigrated to Israel during the early fifties from North Africa, Iraq, and other Mid-eastern countries. Their variables included:

X1 - Ethnicity (*sphrd*) - a dummy variable coded 1 if the respondent or both his parents were born in an Asian or North African country, 0 otherwise

X2 - Parental Education (*pared*) - A scale which ranges from a low of 0 to a high of 1.697

X3 - Scholastic Aptitude (*aptd*) - A composite score based on seven achievement tests.

Y - Grades (*grades*) - Respondent's grade-point average during the first trimester of eighth grade. This scale ranges from a low of 4 to a high of 10.

Analyze these data using Stata and answer the following questions. Begin with the command

regress grades sphrd pared aptd

and then execute whatever other commands are necessary. Note that NO hand computation is needed. All you have to do is run the analyses in Stata and then interpret the results.

1. What is the metric (unstandardized) coefficient for the effect of *aptd* on *grades*? What is the 99% confidence interval for this effect?
2. What is the standardized coefficient for the effect of *sphrd* on *grades*? What is the tolerance of *sphrd*?
3. Test the hypothesis $\beta_{pared} = \beta_{aptd} = 0$. (Remember, it is very easy to do this in Stata.)
4. What percentage of the respondents are Sephardim?
5. What are the partial, semipartial, and zero-order (i.e. bivariate) correlations of *pared* with *grades*?
6. In their published analyses, Shavit and Williams reported that

$$E(\text{grades}) = .185 * \text{sphrd} - .119 * \text{pared} + .49 * \text{aptd} + 4.057$$

Your results should be close, but not identical to this. Explain what might account for the discrepancy.