

2009 Methods Comprehensive Examination  
Department of Sociology, University of Colorado

Question #1: Research Methods (required of all students)

One of the key ideas that often arises in discussions of research methods is the concept of “randomness”. We randomly sample, we randomly assign, we explore the degree to which observed effects are random (vs. fixed), etc. What does the concept of “randomness” actually mean in these situations, why do we give it so much attention, and how is it utilized in surveys, experiments, analyses of existing data, ethnographic research, and meta analyses? How is the concept of randomness related to other key concepts such as “causality”, “heterogeneity”, “manipulation” and “control”? What types of approaches do we draw on to approximate randomness when it is not possible or practical?

Please choose **2 of the following 3** research tasks and write a short description of how you would use the *idea* of randomness to gain empirical leverage on the question at hand.

- A. Develop a survey instrument to measure the relationship between electronic networking (myspace, facebook, etc) and binge drinking among college students.
- B. Use magazine stories about rock stars and album/cd/MP3 sales figures to do a content analysis concerning the relationship between the presentation of rock star sexuality and their financial success over time.
- C. Build a regression model to identify what influences success on graduate methods exams.

2009 Methods Comprehensive Examination  
Department of Sociology, University of Colorado

Question #2: Interpretation of statistical output (required of all students)

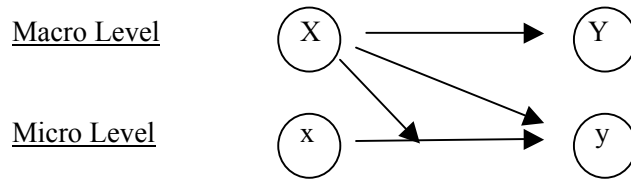
The attached output (at the end of the exam) is taken from a series of runs from the 2006 General Social Survey. The first table is a descriptive table of the dependent variable (yearly income) and the independent variables (gender, race, and education). The independent variables are coded  $dfemale = 1$  (female) and 0 (male);  $age$  (continuous in years);  $dblack = 1$  (self reported black)  $dother = 1$  (self-reported race = other) and the reference category is self-reported race = white;  $educ$  (continuous in years of total education). The next table is simply tabular output for the dependent variable with its corresponding categories; in the first table income ranges from 1 to 25 and these values correspond to the categories listed below.

- 1) Please interpret the meaning of the value for the mean income ( $rincom06$ ) of 14.48.
- 2) Please interpret the meaning of the values of the means for the variables 'dblack' and 'dother'.
- 3) Based on the estimates in Tables 1 and 2, does income appear to be normally distributed, positively skewed, or negatively skewed (please be sure to explain why)?
- 4) Please estimate a 95% confidence interval for the variable  $educ$  (years of education) and interpret the value of this estimate.
- 5) The values in Table 3 are taken from an OLS regression model in which yearly income is regressed on gender, age, and race. Based on the results in Table 3, please answer the following questions:
  - a. What is the meaning of the coefficient for the variable 'dfemale'?
  - b. Estimate a 95% confidence interval for this estimate.
  - c. What is the meaning of the estimate 'beta' for the variable 'dfemale'?
  - d. What is the meaning of the 'R-squared' value?
  - e. Interpret the effects for 'dblack' and 'dother'.
  - f. Interpret the effect of age.
  - g. Interpret the value of the intercept ( $_cons$ ).
- 6) In Table 4, a control for education is included.
  - a. Interpret the difference in the race coefficients from Table 3 to Table 4.
  - b. Does education significantly improve the overall model? Be sure to include the specific information that you are using to make this decision.
  - c. According to the results in Table 4, what variable has the strongest effect on income? Be sure to indicate what information you are using to make this decision.
- 7) In Tables 5 and 6, the researcher examines non-linear specifications of the age effect. In Table 5, she includes a quadratic and in Table 6 she tries a logarithmic transformation.
  - a. Please explain what these coefficients tell you about the association between age and income.
  - b. If you could only present one of the three models (age, age squared, or  $\ln(\text{age})$ ) which model would you present and why?
  - c. Please provide a rough sketch describing estimated yearly income as a function of age based on the quadratic ( $\text{age}^2$ ) results.
- 8) In Table 7, the researcher includes an interaction between race and education ( $ed\_blk = educ*dblack$ ;  $ed\_oth = educ*dother$ ).
  - a. In your own words, describe what is meant by the coefficient  $b = .3637$  (for  $ed\_blk$ ).
  - b. Based on these estimates, please provide a rough sketch describing the effects of education on yearly income by race (white, black, and other).

2009 Methods Comprehensive Examination  
Department of Sociology, University of Colorado

Question 3. Please answer ONE of the following:

3A. The model below represents a group of individual, aggregate, and contextual effects.



Please discuss in detail what each arrow is suggesting. Then propose a substantive research project that could be modeled with the diagram above. Discuss what the proposed relationships would represent and what kind of data would be used to test the various relationships. While you do not need to discuss exactly what methods you would use to analyze the data, strong answers will indicate what some of the important issues concerning data analysis would be.

3B. Statistical significance is a widely used term but it has a very precise meaning. Based on the results from Table 4, please answer the following questions. What exactly is the point of the t-test for the regression coefficient 'educ' and what does the value of 16.75 tell us? Please be sure to elaborate upon the logic of this test (of significance) making specific reference to the parameter estimate, sampling distributions, standard errors, and probabilities. Also, be sure to describe how characteristics of the sample (e.g., sample size, sampling design, etc.) may influence these results. Finally, criticize the use of statistical significance to summarize the size or impact of education on yearly income.

2009 Methods Comprehensive Examination  
 Department of Sociology, University of Colorado

[TABLES FOR QUESTION #2]

TABLE 1

. sum rincom06 dfemale age age2 dblack dother educ

Variable	Obs	Mean	Std. Dev.	Min	Max
rincom06	2656	14.47666	5.742581	1	25
dfemale	2656	.5011295	.5000929	0	1
age	2656	42.14533	12.96505	18	88
dblack	2656	.1340361	.3407553	0	1
dother	2656	.1411898	.3482828	0	1
educ	2656	13.73456	3.11039	0	20

TABLE 2

. tab rincom06

**The lowest income (< \$1k) is coded 1 and the highest income (>\$150k) = 25**

respondents income	Freq.	Percent	Cum.
under \$1 000	52	1.96	1.96
\$1 000 to 2 999	75	2.82	4.78
\$3 000 to 3 999	47	1.77	6.55
\$4 000 to 4 999	52	1.96	8.51
\$5 000 to 5 999	41	1.54	10.05
\$6 000 to 6 999	42	1.58	11.63
\$7 000 to 7 999	37	1.39	13.03
\$8 000 to 9 999	61	2.30	15.32
\$10000 to 12499	122	4.59	19.92
\$12500 to 14999	108	4.07	23.98
\$15000 to 17499	110	4.14	28.13
\$17500 to 19999	111	4.18	32.30
\$20000 to 22499	138	5.20	37.50
\$22500 to 24999	108	4.07	41.57
\$25000 to 29999	192	7.23	48.80
\$30000 to 34999	221	8.32	57.12
\$35000 to 39999	167	6.29	63.40
\$40000 to 49999	299	11.26	74.66
\$50000 to 59999	198	7.45	82.12
\$60000 to 74999	175	6.59	88.70
\$75000 to \$89999	104	3.92	92.62
\$90000 to \$109999	64	2.41	95.03
\$110000 to \$129999	50	1.88	96.91
\$130000 to \$149999	22	0.83	97.74
\$150000 or over	60	2.26	100.00
Total	2,656	100.00	

2009 Methods Comprehensive Examination  
 Department of Sociology, University of Colorado

TABLE 3  
 . reg rincom06 dfemale age dblack dother, beta

Source	SS	df	MS	Number of obs = 2656	
Model	8908.83225	4	2227.20806	F( 4, 2651)	= 75.08
Residual	78645.7205	2651	29.6664355	Prob > F	= 0.0000
-----				R-squared	= 0.1018
-----				Adj R-squared	= 0.1004
Total	87554.5527	2655	32.9772327	Root MSE	= 5.4467

rincom06	Coef.	Std. Err.	t	P> t	Beta
dfemale	-2.694859	.2121302	-12.70	0.000	-.2346819
age	.0805726	.0082526	9.76	0.000	.1819092
dblack	-.7539923	.3160376	-2.39	0.017	-.0447407
dother	-1.404363	.3108626	-4.52	0.000	-.0851735
_cons	12.73071	.392638	32.42	0.000	.

TABLE 4  
 . reg rincom06 dfemale age dblack dother educ, beta

Source	SS	df	MS	Number of obs = 2656	
Model	16436.5116	5	3287.30232	F( 5, 2650)	= 122.49
Residual	71118.0411	2650	26.8369966	Prob > F	= 0.0000
-----				R-squared	= 0.1877
-----				Adj R-squared	= 0.1862
Total	87554.5527	2655	32.9772327	Root MSE	= 5.1804

rincom06	Coef.	Std. Err.	t	P> t	Beta
dfemale	-2.829823	.2019216	-14.01	0.000	-.2464353
age	.0793034	.0078496	10.10	0.000	.1790436
dblack	-.4828547	.3010246	-1.60	0.109	-.0286518
dother	-.3267787	.3025866	-1.08	0.280	-.0198189
educ	.5551036	.0331444	16.75	0.000	.3006643
_cons	5.03925	.591919	8.51	0.000	.

TABLE 5  
 . reg rincom06 dfemale age age2 dblack dother educ, beta

Source	SS	df	MS	Number of obs = 2656	
Model	24253.4963	6	4042.24938	F( 6, 2649)	= 169.16
Residual	63301.0565	2649	23.8962086	Prob > F	= 0.0000
-----				R-squared	= 0.2770
-----				Adj R-squared	= 0.2754
Total	87554.5527	2655	32.9772327	Root MSE	= 4.8884

rincom06	Coef.	Std. Err.	t	P> t	Beta
dfemale	-2.871318	.1905513	-15.07	0.000	-.2500488
age	.8265805	.0419755	19.69	0.000	1.866175
age2	-.0085288	.0004716	-18.09	0.000	-1.713581
dblack	-.6476974	.2841993	-2.28	0.023	-.0384333
dother	-.3823774	.2855436	-1.34	0.181	-.0231909
educ	.512485	.0313644	16.34	0.000	.2775805
_cons	-9.236666	.9669482	-9.55	0.000	.

2009 Methods Comprehensive Examination  
 Department of Sociology, University of Colorado

TABLE 6

. reg rincom06 dfemale ln\_age dblack dother educ, beta

Source	SS	df	MS		
Model	18302.779	5	3660.55579	Number of obs =	2656
Residual	69251.7738	2650	26.1327448	F( 5, 2650) =	140.08
				Prob > F =	0.0000
				R-squared =	0.2090
				Adj R-squared =	0.2076
				Root MSE =	5.112
Total	87554.5527	2655	32.9772327		

rincom06	Coef.	Std. Err.	t	P> t	Beta
dfemale	-2.833803	.1992539	-14.22	0.000	-.2467818
ln_age	4.106786	.3093557	13.28	0.000	.2320525
dblack	-.4473569	.2968814	-1.51	0.132	-.0265454
dother	-.2239625	.2983345	-0.75	0.453	-.0135831
educ	.5471031	.0327161	16.72	0.000	.2963309
_cons	-6.683216	1.235979	-5.41	0.000	.

TABLE 7

. reg rincom06 dfemale age age2 dblack dother educ ed\_blk ed\_oth, beta

Source	SS	df	MS		
Model	24548.0195	8	3068.50244	Number of obs =	2656
Residual	63006.5332	2647	23.802997	F( 8, 2647) =	128.91
				Prob > F =	0.0000
				R-squared =	0.2804
				Adj R-squared =	0.2782
				Root MSE =	4.8788
Total	87554.5527	2655	32.9772327		

rincom06	Coef.	Std. Err.	t	P> t	Beta
dfemale	-2.888883	.1902652	-15.18	0.000	-.2515785
age	.8209913	.0419436	19.57	0.000	1.853556
age2	-.0084749	.000471	-17.99	0.000	-1.702752
dblack	-5.607012	1.538979	-3.64	0.000	-.3327109
dother	.0810189	.9460399	0.09	0.932	.0049137
educ	.4915505	.0393866	12.48	0.000	.2662416
ed_blk	.3637876	.1107261	3.29	0.001	.2993929
ed_oth	-.0421103	.0710886	-0.59	0.554	-.0330883
_cons	-8.80173	1.016341	-8.66	0.000	.